Jetstream Stakeholder Advisory Board Meeting February 2017: Presenters' Report

Craig A. Stewart[1], *PI*
David Y. Hancock[1], *Systems Lead*
Matthew Vaughn[2], *Co-PI*
Nirav Merchant[3], *Co-PI*
J. Michael Lowe[1], *Jetstream lead sysadmin*
Jeremy Fischer[1], *Senior Technical Advisor*
Lee Liming[4], *Globus Services Lead and SI*
James Taylor[5], *Jetstream Co-PI and Galaxy PI*
George Turner[1], *System Architect*
C. Bret Hammond[1], *Jetstream sysadmin*
Edwin Skidmore[3], *Atmosphere software lead*
Michael Packard[2], *Senior systems administrator*
Therese Miller[1], *Project manager*
Paul Rad[6], *Jetstream collaborator*
Susan Mehringer[7], *Jetstream collaborator*
Ian Foster[4], *Co-PI*

[1]Indiana University Pervasive Technology Institute
[2]University of Texas at Austin Texas Advanced Computing Center
[3]University of Arizona
[4]University of Chicago Computation Institute
[5]Johns Hopkins University
[6]University of Texas, San Antonio
[7]Cornell University

February 28, 2017

Stewart, C.A., Hancock, D.Y., Vaughn, M., Merchant, N., Lowe, J.M., Fischer, J., Liming, L., Taylor, J., Hammond, C.B., Skidmore, E., Miller, T., Rad, P., Mehringer, S. & Foster, I. (2017). Jetstream Stakeholder Advisory Board Meeting February 2017: Presenters' Report (PTI Technical Report PTI-TR17-001) Bloomington, IN: Indiana University. Retrieved from http://hdl.handle.net/2022/21247

Table of Contents

1. Introduction

As a part of the National Science Foundation (NSF) Jetstream project (award #1445604), a Stakeholder Advisory Board (SAB) Meeting was held February 6, 2017 – February 7, 2016.

In our proposal to the NSF, which subsequently resulted in NSF award 1445604, we proposed the Jetstream system which provides:

- "Self-serve" academic cloud services, enabling researchers or students to select a VM image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing. Authentication to this "self-serve" environment is via Globus using XSEDE credentials.
- Hosting for persistent Science Gateways. Jetstream supports persistent science gateways, including the capability of hosting persistent science gateways within a VM when the nature of the gateway is consistent with operation within a VM. o Galaxy is one of the initial science gateways supported.
- Data movement, storage and dissemination.
    - Jetstream supports data transfer with Globus Connect.
    - Users are able to store VMs in the Indiana University persistent digital repository, IUScholarWorks (scholarworks.iu.edu) and obtain a Digital Object Identifier (DOI) that is associated with the VM stored.
- Virtual Linux desktop services delivered from Jetstream to tablet devices. This service is aimed at increasing access to Jetstream for users at institutions with limited resources including small schools, schools in EPSCoR states, and Minority Serving Institutions.

In this document, we present the agenda for the meeting and the presentations given to the SAB by the Jetstream team members.

2. Meeting agenda

08:30 Introductions and logistics; with some background from each participant

Charge for the SAB:

- Offer advice and assessment on our plans for meeting guidance from the NSF review panel on things we should be doing, but did not have in the original proposal
- Offer assessment on our progress (and presentation of progress) against milestones in the original proposal
- What should we be thinking about in the long run for success in the remainder of this project (and we write this thinking that we can run through the rest of the project and do a good job of hitting most of the agreed upon metrics - what do we do to get the most value out of the system)
    - In particular what does the SAB think of our priorities and plans for supplement requests
- How can we aid computer and computational science research (e.g. how do we create more linkages with Chameleon, CloudLab, etc and what can we do to aid their future sustainability)
- What should we be thinking about beyond the original context of Jetstream

09:00 The case as we make it now for Jetstream (with comments about the original case in the proposal) (Stewart)

- What does the SAB think the case ought to be?
- Deviation from plan: API use higher relative to Atmosphere

09:30 Usage Statistics

10:00 Break

10:15 Science highlights, survey results, and new projects being brought onboard now

- Does the SAB have any ideas about high importance / high priority projects we ought to contact regarding use of Jetstream

10:45 Project plan for next year

- Atmosphere interface and software plans (Edwin Skidmore)
- Accounting process plans (Matt Vaughn)
- Globus services (Lee Liming)
- Science Gateway plans (Marlon Pierce)
    - State of gateways on Jetstream
    - Future
        - Alignment with SGCI
        - Part of the XSEDE Gateway Hosting solution, replacing Quarry
- New Hardware (Matt Vaughn & Mike Lowe)

12:00 Lunch brought in; discussion of outreach plans for the year
Specific topic: how do we attract more engineering users? (Therese Miller)

1:00 Break

1:15 Discussions

- What should we do going forward - a few specific discussion starters
- XSEDE PIF requests (Project Improvement Fund) (by the way we are paying attention to project Aristotle)
- Current priorities for supplement requests
- Any suggestions on a CISO to add to the SAB?
- Discussion of CPU % utilization as a metric; are there places we can get numbers for comparisons?
- Discussion of Jetstream stats as possible resources for research
- Futures:
  - Short term: AWS, Google, Azure current and future features we should pay attention to and try to emulate
  - Mid term: schedulers?
  - Long term: IoT, fog computing, "serverless" computing, real-time "edge computing."

3:00 Wrap up

3. Presentations by Jetstream Team Members

List of presentations:

1. Craig A. Stewart, "Jetstream Overview for Jetstream Stakeholder Advisory Board"
2. George Turner, "Jetstream Metrics"
3. Craig A. Stewart, "A sampling of important projects – past and future"
4. Craig A. Stewart, "Jetstream User Survey 2016 – Preliminary Results"
5. Edwin Skidmore, "CyVerse Atmosphere"
6. Matthew Vaughn and J. Michael Lowe, "Jetstream Software Roadmap"
7. Lee Liming, "Globus features"
8. Marlon Pierce, "Jetstream Support for Science Gateways"
9. Craig A. Stewart, "Looking forward to new technology"
10. Therese Miller, Susan Mehringer, Paul Rad, "Outreach, Education and Training Plans for 2017"

Jetstream Overview for Jetstream Stakeholder Advisory Board

Craig A. Stewart, Jetstream PI
7 February 2016

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

Cyberinfrastructure

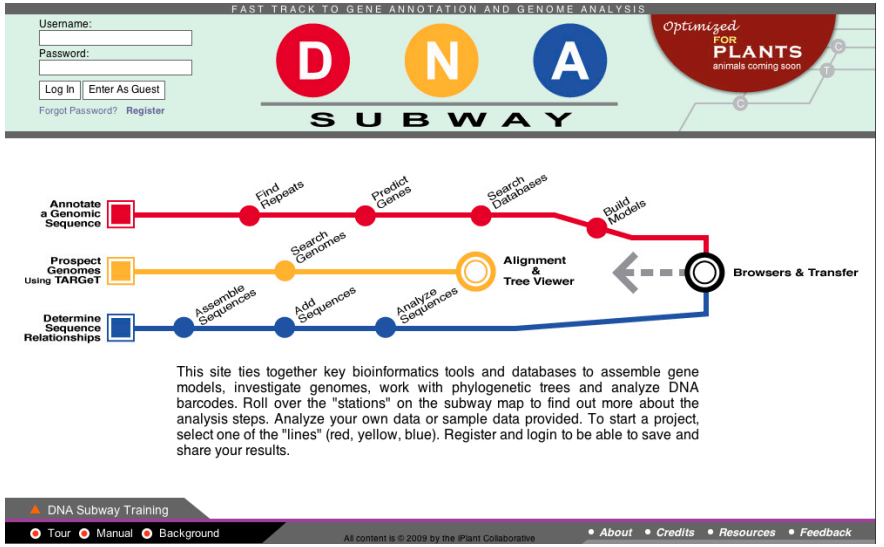
- Gained wide usage with 2003 "Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure" (aka the Atkins Report)
- Definition we use: *Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to* ***improve research productivity and enable breakthroughs not otherwise possible.***
- *Question: Does anyone on the SAB want or need an explanation of the relationship between Jetstream and XSEDE*

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Recent major biologically-oriented grant awards by the NSF for cyberinfrastructure in the form of software

- iPlant (now called CyVerse)
- Galaxy



led by the National Science Foundation
Award #ACI-1445604

What about NSF-funded hardware impact across research communities?

- Around 350,000 researchers, educators, & learners received NSF support in 2015
  - <2% completed a computation, data analysis, or visualization task on XD/XSEDE resources
  - 70% of researchers surveyed* claimed to be resource constrained
- Why are the people not using NSF-funded supercomputers not using them?
  - Perceived ease of access and use
  - HPC resources –are often not well-matched to their needs, and they just don't need that much capability
- So the NSF put out a proposal solicitation for CI systems to support a broader spectrum of NSF-funded researchers

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

The four major functions we identified for Jetstream

- "Self-serve" academic cloud services, enabling researchers or students to select a VM image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing.
- Hosting of persistent VMs to provide services beyond the command line interface for science gateways and other science services.
- Enable data movement, storage, and dissemination
  - Jetstream supports data transfer with Globus tools.
  - Users are able to store VMs in the Indiana University digital repository,IUScholarWorks, and make them discoverable with a Digital Object Identifier (DOI).
- Provide virtual desktop services to tablet devices, increasing CI access for users at resource-limited institutions (e.g., small schools, schools in EPSCoR states, and Minority Serving Institutions), thus expanding access to Jetstream and the NSF XSEDE-supported ecosystem (XSEDE is the eXtreme Science and Engineering Discovery Environment).

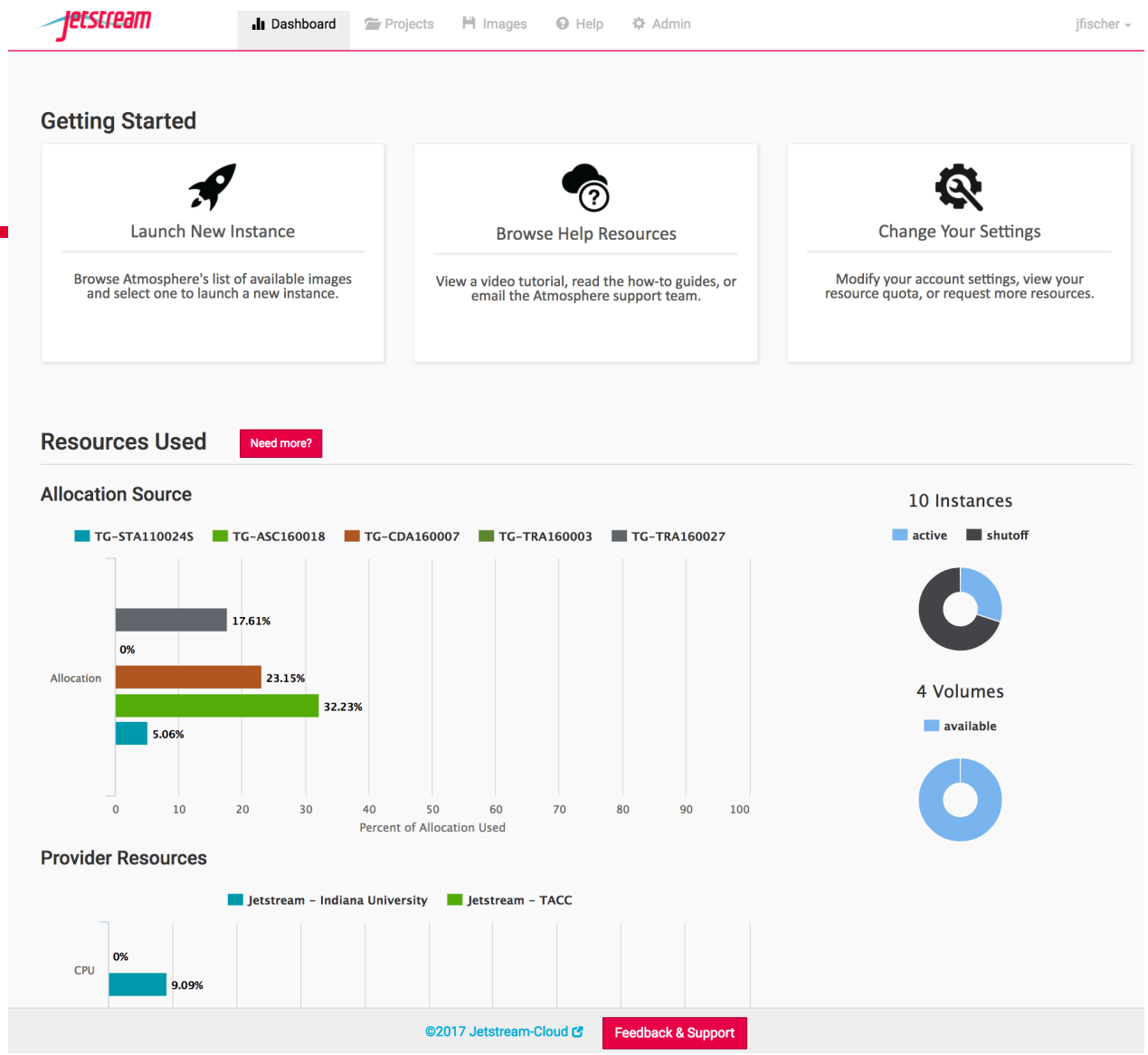
Jetstream
http://jetstream-cloud.org/

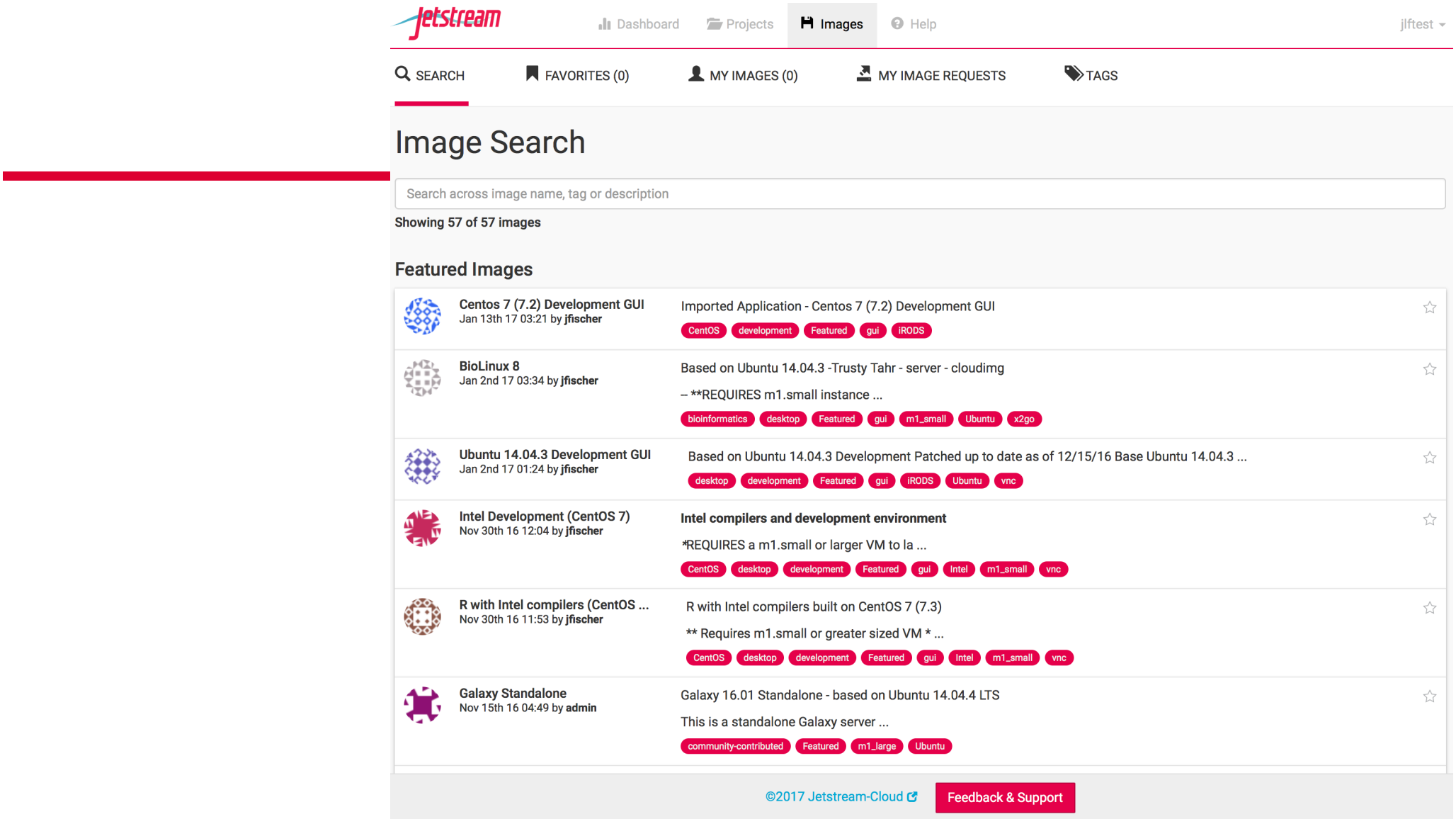
NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream responds to many the research communities' requests:

- What do we want?
  - to be able to do our research
- When do we want it?
  - When we want it / when we need it
- And what else?
  - We want to be able to disseminate our analyses and make them easily replicable, so that when we find interesting and important results people believe us!

Jetstream web interface




Jetstream
Dashboard
Projects
Images
Help
jlftest
SEARCH
FAVORITES (0)
MY IMAGES (0)
MY IMAGE REQUESTS
TAGS
Image Search
Search across image name, tag or description
Showing 57 of 57 images
Featured Images
Centos 7 (7.2) Development GUI
Jan 13th 17 03:21 by jfischer
Imported Application - Centos 7 (7.2) Development GUI
CentOS development Featured gui iRODS
BioLinux 8
Jan 2nd 17 03:34 by jfischer
Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg
-- **REQUIRES m1.small instance ...
bioinformatics desktop Featured gui m1_small Ubuntu x2go
Ubuntu 14.04.3 Development GUI
Jan 2nd 17 01:24 by jfischer
Based on Ubuntu 14.04.3 Development Patched up to date as of 12/15/16 Base Ubuntu 14.04.3 ...
desktop development Featured gui iRODS Ubuntu vnc
Intel Development (CentOS 7)
Nov 30th 16 12:04 by jfischer
Intel compilers and development environment
*REQUIRES a m1.small or larger VM to la ...
CentOS desktop development Featured gui Intel m1_small vnc
R with Intel compilers (CentOS ...
Nov 30th 16 11:53 by jfischer
R with Intel compilers built on CentOS 7 (7.3)
** Requires m1.small or greater sized VM * ...
CentOS desktop development Featured gui Intel m1_small vnc
Galaxy Standalone
Nov 15th 16 04:49 by admin
Galaxy 16.01 Standalone - based on Ubuntu 14.04.4 LTS
This is a standalone Galaxy server ...
community-contributed Featured m1_large Ubuntu
©2017 Jetstream-Cloud
Feedback & Support

Jetstream is a managed science cloud…

… a cloud managed for science

Major target disciplines:

- Life Sciences (biology other than protein folders)
- Earth sciences
- Engineering
- Sociology
- Economics
- Observational astronomy
- Outreach across underrepresented communities

Community partners

Discipline or mode of use	Lead partners
Biology	iPlant, University of Arizona*, Galaxy, Johns Hopkins University*, National Genome Analysis Support Center
Earth Science/Polar Science	National Snow and Ice Data Center Network (RCN)
Field station research	University of Arizona
Network Science	IU Network Institute
Observational astronomy	WIYN Consortium
Social sciences	Odum Institute, University of North Carolina
Campus bridging	XSEDE, Cornell, IU
Under-resourced schools	University of Hawaii, Jackson State University, University of Texas San Antonio*, resetting with U. Arkansas Pine Bluff
Use of proprietary software	Mathworks
Reproducible data analyses	University of Chicago Computation Institute
Enhanced science gateway deployment	University of California San Diego (Supercomputing Center), XSEDE
Visualization and analysis	IU, University of Texas

Keys to what we do

The Jetstream team manages Jetstream from top to bottom so as to respond to user needs.

We treat every user the way AWS treats their large customers

The team incorporates everyone from software developers to users; the communities we aim to serve are integrated into the team

The Atmosphere interface, Globus tools, Apache Arivatha supporting Gateways

How are we doing on recommendations from the review panel?

- Share perspectives and experiences gained … as contributions towards potential future NSF consideration of NSF system acquisition evaluative criteria. **Done.**
- Add an ISSO to the Jetstream advisory committee. **Not done. A recommendation from the SAB would be welcome.**
- Develop a Secure Software MarketPlace, including perhaps in collaboration with ongoing synergistic developments @ the IU Center for Applied Cybersecurity Research, including with Von Welch, to further accelerate advances in "trusted virtual machines" *and* Accelerate exploration of deploying trusted containers as a contributions further advancing system level security controls. **Underway.**
- Expand the composition of the advisory board to include representatives of engineering communities. **Minimal progress made; advice and assistance would be appreciated**
- Further accelerate engagements with science and engineering research communities contributing to prioritize science and engineering applications, potentially through science gateways, including as contribution towards maintaining project focus and scope. **Some progress made, but we feel we are not on track as regards engineering applications**
- Accelerate scalable outreach through students, including potentially those who have successfully used Jetstream including to perhaps produce a YouTube video as a potential contribution towards significantly broadening awareness of Jetstream, including perhaps if the resulting video were to "go viral." **Some tiny bits of progress made: https://ensemble.brandeis.edu/Watch/a9Z5Ypi2**
- Bonus goal: what are you doing for field stations. **Good progress underway.**

What we have learned so far

When we wrote a definition for cyberinfrastructure and the line "*improve research productivity and enable breakthroughs not otherwise possible*" we were not thinking of Jetstream. But is has certainly done that.

Our biggest mistake: Agreeing to a target figure for CPU utilization and then agreeing to 6% as that figure

XSEDE's biggest mistake: Not having a process for onboarding new communities

Biggest success: field biology

Biggest surprise: interest in Atmosphere API use

Biggest unsolved problem: attracting engineers to use Jetstream

Biggest news working with NSF: Bob Chadduck is a wonderful Program Officer

A few thoughts to start off the day

This is an unusual group: it IS a Stakeholder Advisory Board, and we want to spend most of our time listening to you

Gwen Jacobs will keep us on track and summarize areas of consensus and areas of variety of opinion as we go

I would like to request a very concise report from the committee if possible (1-2 pages)

Jetstream Partner Organizations

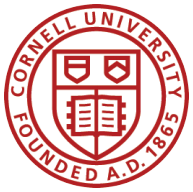
Initial construction (funded partners)

THE UNIVERSITY OF ARIZONA

THE UNIVERSITY OF CHICAGO

JOHNS HOPKINS UNIVERSITY

Management & Operations partners



Application / community lead partners

UNC THE ODUM INSTITUTE

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Questions and discussion?

Stakeholder's Advisory Board Meeting
Chicago, IL 7-Feb-2017

Jetstream Metrics

George Turner
Indiana University
Research Systems Architect

funded by the National Science Foundation
Award #ACI-1445604

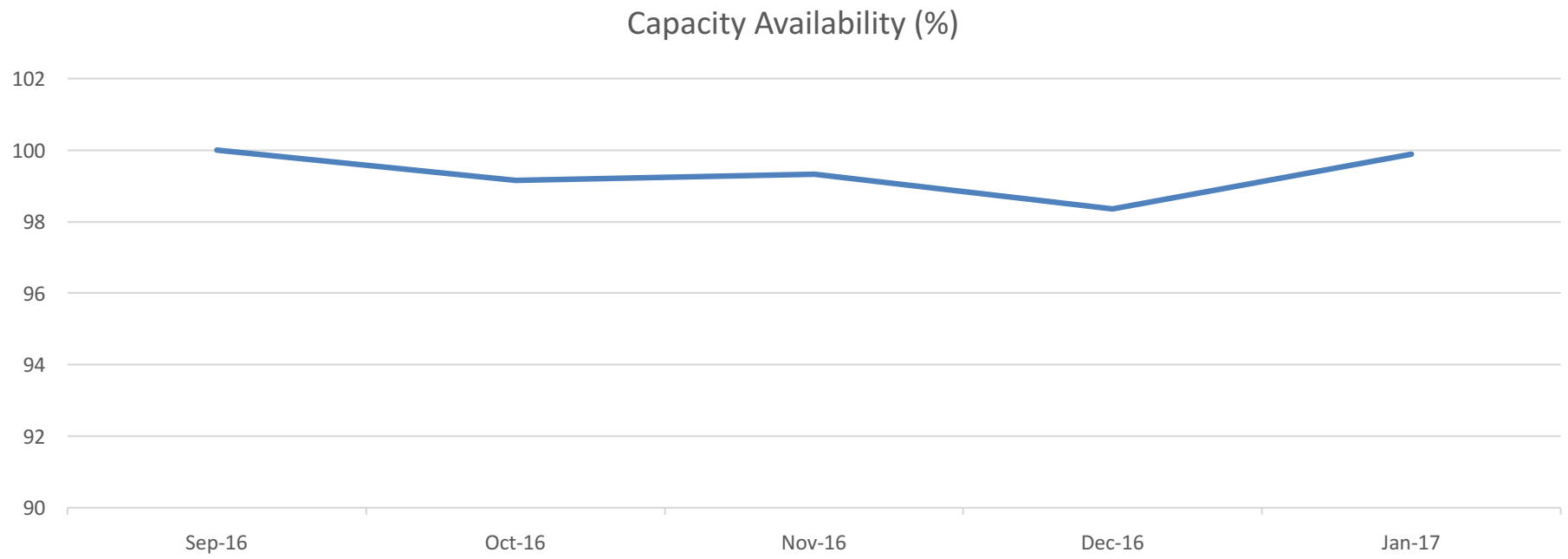
Operational Metrics Overview

	Jun 2016	Jul 2016	Aug 2016	PY Q1 Total	Sep 2016	Oct 2016	Nov 2016	PY Q2	Dec 2016	Jan 2017	Goal	
System availability	*	*	*	99.29%	100%	100%	100%	100%	100%	100%	95%	G
Capacity availability	*	*	*	99.65%	100%	98.88%	98.54%	99.14%	98.36%	99.88	95%	G
Job completion success	*	*	*	99.01%	97.83%	99.66%	100%	99.16%	98.09%	98.37%	96%	G
Core cloud environment software release	Liberty	Mitaka	Mitaka	N/A	Mitaka	Mitaka	N/A	Mitaka	Mitaka	Mitaka	Maintain Currency	G
Average number of active VMs	*	160 IUonly	211 IUonly	186 IUonly	267	379	380	342	517	584	320	G
Peak active VMs	*	336 IUonly	336 IUonly	540	348	560	551	486	610	653	320	G
CPU % average utilization	*	*	*	0.60%	*	1.05% IUonly	*	0.74	*	*	6%	R
CPU % peak	*	*	*	2.34%	*	1.52% IUonly	*	4.39%	*	*	6%	R

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Capacity Availability (Goal 95%)



Jetstream
http://jetstream-cloud.org/

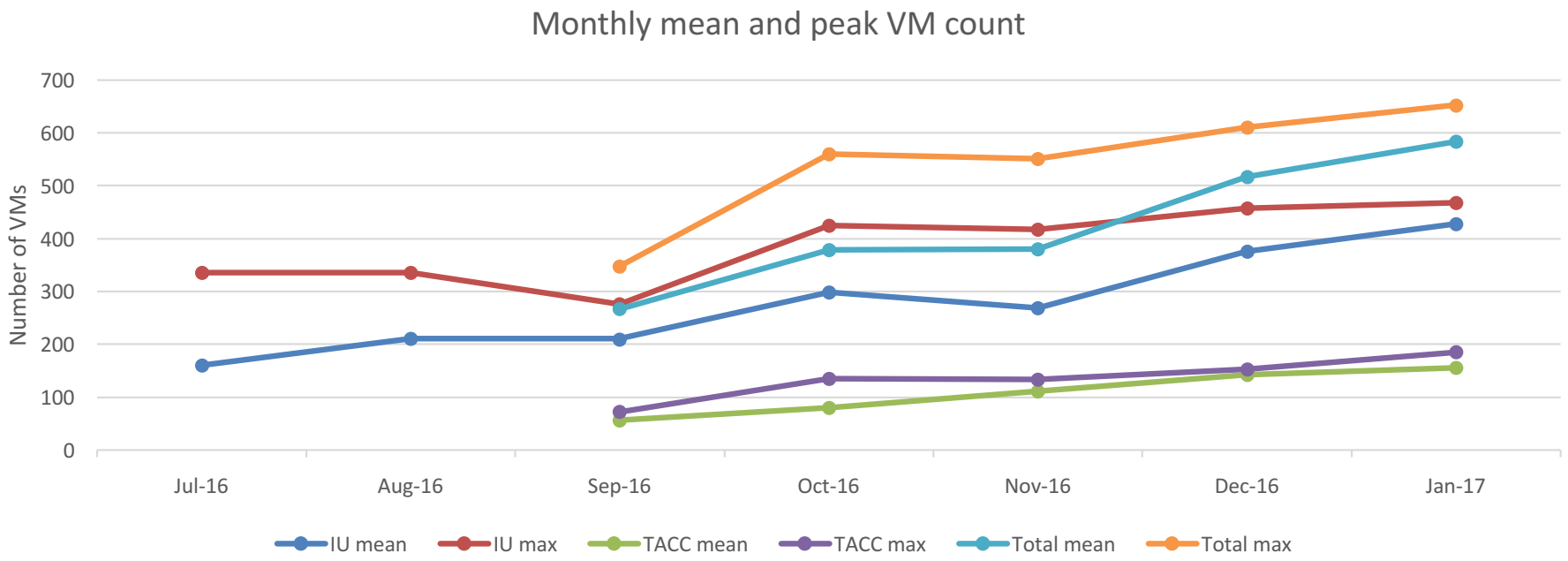
funded by the National Science Foundation
Award #ACI-1445604

% Job Completion Success (96%)

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Average, Peak number of active VMs

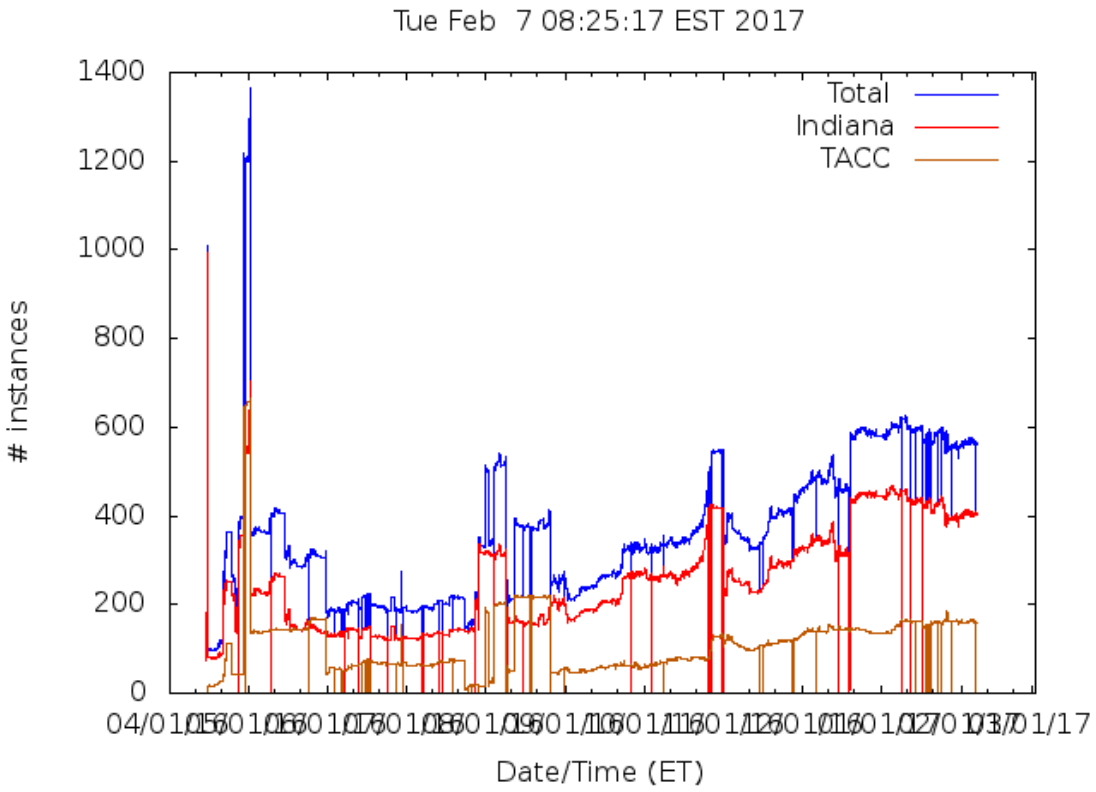


Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Number of "Active" VMs since 14-Apr-2016

http://mypage.iu.edu/~turnerg/plot-instances-all.png

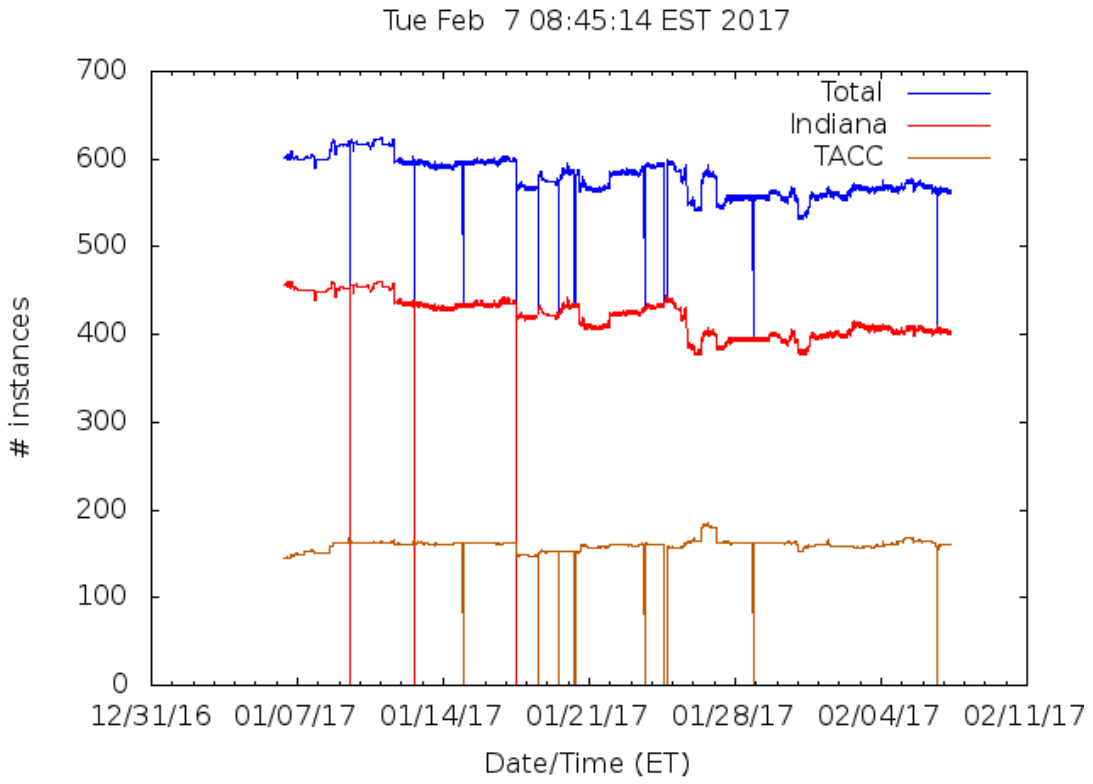


Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Number of "Active" VMs since 14-Apr-2016

http://mypage.iu.edu/~turnerg/plot-instances-month.png



Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

SUs, Users,

	Q3 Jun - Aug	Q4 Sep - Nov	Dec 2016	Annual running total	Status	Targets for metrics and Notes
Capacity of system allocated via NSF-specified allocation process	96.5%	>100%	>100%	98.25%	G	90%
Total number of distinct users	593	1,463	1,547	1,463	G	1,000; annual goal already exceeded
Total number of students having used Jetstream in an educational or training setting	36	281	336	281	G	100; annual goal already exceeded
Total number of science gateways using Jetstream (running total)	5	5	9	9	G	2; annual goal already exceeded
SUs available to user community per month	4.41M	4.41M	1.47M	10.29M	G	1.47M (per month)
% of SUs available to user community that were used per month	--	79.81%	>100% (2227411 .54 SUs)	89.91%	W	--

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Products

	Q3 Jun - Aug	Q4 Sep - Nov	Dec 2016	Annual running total	Status	Targets for metrics and Notes
Total number of publications facilitated by use of Jetstream	2	2	4	8		5
Total number of VM images and/or data sets published with a DOI via IUScholarWorks	9	1	1	11		10; annual goal already met

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Allocations, total

	PY Q1	Sept	October	November	PY Q2	December	PY Total Year to Date
Total request and allocations							
Total requests	67	31	45	11	87	15	169
Total SUs requested	3,804,622	2,280,000	18,455,736	700,000	21,435,736	975,000	26,215,358
Total SUs awarded	3,804,622	2,280,000	16,955,736	700,000	19,935,736	975,000	24,715,358

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Allocations, Startups & Educational

	PY Q1	Sept	October	November	PY Q2	December	PY Total Year to Date
Startup							
Total requests	13	12	19	7	38	5	56
SUs requested	810,062	1,130,000	1,719,000	500,000	3,349,000	250,000	4,409,062
SUs awarded	810,062	1,130,000	1,719,000	500,000	3,349,000	250,000	4,409,062
Educational							
Total requests	6	1	0	0	1	2	9
SUs requested	392,000	50,000	0	0	50,000	125,000	567,000
SUs awarded	392,000	50,000	0	0	50,000	125,000	567,000

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Allocations, Campus Champions & Supplemental

	PY Q1	Sept	October	November	PY Q2	December	PY Total Year to Date
Campus Champion							
Total requests	45	13	11	4	28	7	80
SUs requested	2,250,000	650,000	550,000	200,000	1,400,000	350,000	4,000,000
SUs awarded	2,250,000	650,000	550,000	200,000	1,400,000	350,000	4,000,000
Supplemental							
Total requests	1	5	1	0	6	1	8
SUs requested	50,000	450,000	1,500,000	0	1,950,000	250,000	2,250,000
SUs awarded	50,000	450,000	1,500,000	0	1,950,000	250,000	2,250,000

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Allocations, Research

	PY Q1	Sept	October	November	PY Q2	December	PY Total Year to Date
Research allocations							
Total requests	2	N/A	14	N/A	14	N/A	16
SUs requested	302,560	N/A	14,686,736	N/A	14,686,736	N/A	14,989,296
SUs awarded	302,560	N/A	13,186,736	N/A	13,186,736	N/A	13,489,296

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Allocations, summary, PY 2016

	Requested	Awarded
Total	26,215,358	24,715,358
Research	14,989,296	13,489,296 (90%)
Supplemental	2,250,000	2,250,000
Startup	4,409,062	4,409,062
Campus Champions	4,000,000	4,000,000
Educational	567,000	567,000

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Statistics requests from the SAB

funded by the National Science Foundation
Award #ACI-1445604

Ratio of API users to Atmosphere users (Jan-2017)

IU API	656317
TACC API	768032
Total API	1424349 (49.5%)
Atmosphere	1450926 (50.5%)
Total SUs	2875275

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Of the total of all the VMs launched how many ran at IU? TACC? How many total Virtual Machines (VMs) were launched?

	Ceilometer	Nova Compute
IU	11,503	17,853
TACC	983	6,590
Total	12,486	24,443

The first column are statistics from Ceilometer which started collecting statistics on or about 1-July and will be kept indefinitely. The second column is from the Nova compute database which has statistics from the beginning; but, at some point will be purged on a regular basis.

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

How big were the largest VMs launched?
How small were the smallest VMs launched?

Flavors: six predefined sizes

Name	Cores	RAM(GB)	Disk(GB)	#/server
Tiny	1	2	8	46
Small	2	4	20	23
Medium	6	16	60	7
Large	10	30	120	4
XLarge	**24**	**60**	**240**	**2**
XXLarge	44	120	480	1

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

What was the average size of VMs launched?

value	count(*)	C/ R/ D
m1.tiny	1650	1/ 2/ 8
m1.small	980	2/ 4/ 20
m1.medium	1179	6/ 16/ 60
m1.large	638	10/ 30/120
m1.xlarge	**6054**	**24/ 60/240**
m1.xxlarge	1002	44/120/480

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

How many XXLs were launched at once?

- Max number at “once”
  - IU 9
  - TACC 2
- Mean number started for each cloud
  - IU 1.69
  - TACC 1.03

Notes: “at once” = within a 10 second window
Only for XXL flavor
Each cloud maintain its own, separate database

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

How long did the **XXL**s instances run?

- IU (hours)
  - Min 0.0022
  - Ave 83.72
  - Max 2227.24
- TACC (hours)
  - Min 0.0033
  - Ave 125.43
  - Max 2073.84

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

How long did the **Tiny** instances run?

- IU (hours)
  - Min 0.0000
  - Ave 129.01
  - Max 7465.79
- TACC (hours)
  - Min 0.0019
  - Ave 171.66
  - Max 6589.03

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

How many Tiny instances were launched at once?

- Max number per cloud
  - IU 5
  - TACC 3
- Mean number per cloud
  - IU 1.06
  - TACC 1.01

Notes: "at once" = within a 10 second window

Each cloud maintains its own, separate databases

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

What was the average number of VMs launched at once?
What was the average run time?

- Mean number launched
  - IU 1.66
  - TACC 1.02
- Mean run time
  - IU 102.25 hours
  - TACC 93.11 hours

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

** What applications ran the most?
** What applications ran on the largest VMs?
** What applications on smallest VMs?

- Do not track applications run within VMs
- We cannot realistically obtain this data
- Intel 7.2 dev is the most launched image
- Users don't change the name of the image they started with which skews the stats

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Who were the top 5-7 users by Project?
What scale of resources did they use?
How does usage map to proposal use cases?

Grant Number	SU Used per use Jan-2017	Board Type	Grant Title	Field of Science	XL hours	XL Days	Concurrent XL
BIO150062	480557	Research	Request for XSEDE allocation on Jetstream for iPlantcollaborative	Biological Sciences	20023.19	834.30	26.9
TRA160003	266281	Startup	Jetstream Test Allocation (research)	Training (Computational fluid dynamics, atmospheric science)	11095.05	462.29	14.9
PHY140033	168792	Research	Renewal Proposal: Cloud Computing on Jetstream for the ATLAS Experiment at the Large Hadron Collider	Elementary Particle Physics	7033.00	293.04	9.5
ASC160018	163816	Startup	Computational Research Analytics Platforms for Problem Investigation	Advanced Scientific Computing	6825.68	284.40	9.2
TRA160003	136880	Startup	Jetstream Test Allocation (outreach & testing)	Training	5703.35	237.64	7.7
MCB140147	114558	Research	The Galaxy XSEDE Gateway	Biological Sciences	4773.25	198.89	6.4
MCB140255	109404	Startup	Computational support for small angle scattering for advanced analyses of structural data in chemical biology and soft condensed matter	Biophysics	4558.50	189.94	6.1

Jetstream
http://jetstream-cloud.org/
funded by the National Science Foundation
Award #ACI-1445604

Can we could share log data for events (VM creation, deletion, etc) asthat's a frequent cloudlab request for people running experiments?

- This is a moving target
  - The dispatcher process can output events to an HTTP URL
    - re: ceilometer.dispatcher.http module
- In the Newton version of OpenStack events are broken out from the Ceilometer project into a new separate service named “Panko”
- It should be possible to offer this once Jetstream is upgraded from Mitaka to Newton

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Status of CPU statistics gathering

- Our statistics gathering is interfering with users' usage.
- Takes an estimated 14 seconds to walk the data structures in the IU cloud.
- Heartbeat is 15 seconds; slightest slow down causes a Ceph OSD to appear offline resulting in a further slowing down.
- Once you get behind, it's an accelerating process.
- We have adjusted parameters and are in the process of catching up.
- Looking to purchase hardware to allow separating out the statistics databases from the databases OpenStack uses for operations.
- Would like to capture more; list of metrics http://docs.openstack.org/admin-guide/telemetry-measurements.html

funded by the National Science Foundation
Award #ACI-1445604

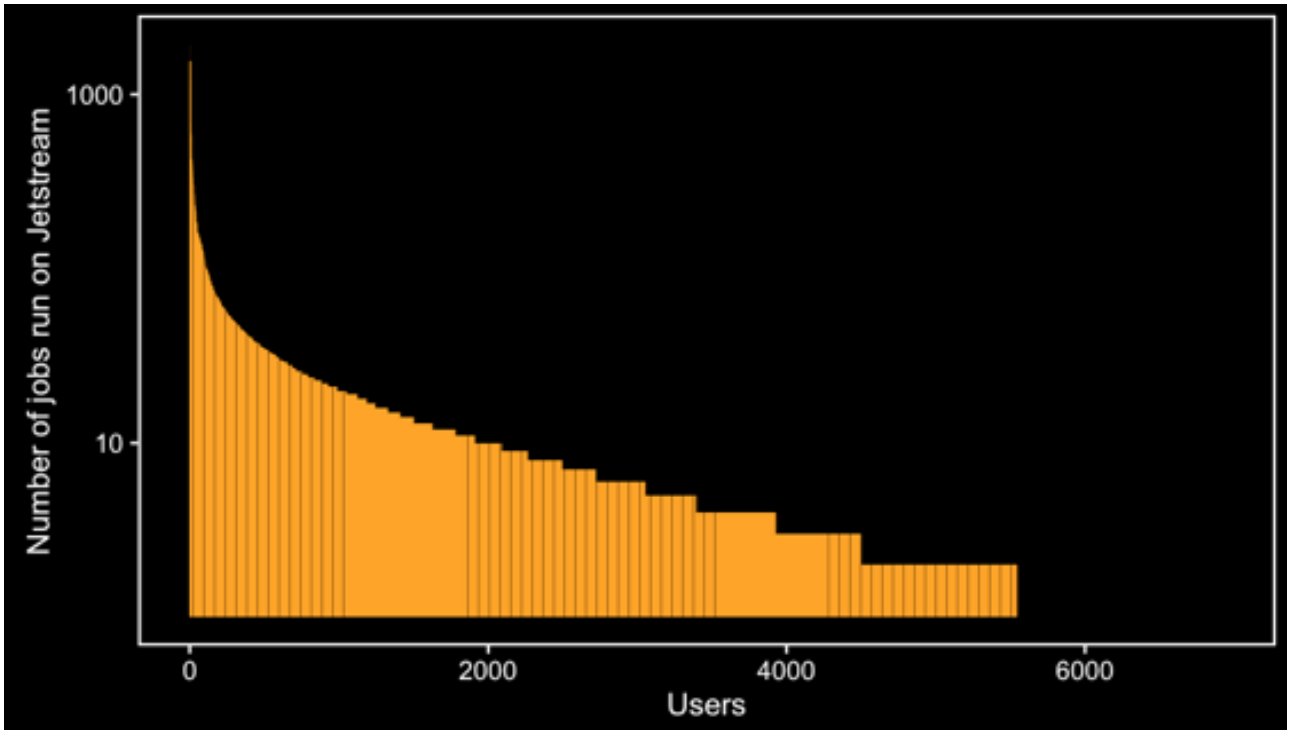
A sampling of important projects – past and future

Craig A. Stewart

Some work already done with Jetstream

Galaxy use of Jetstream

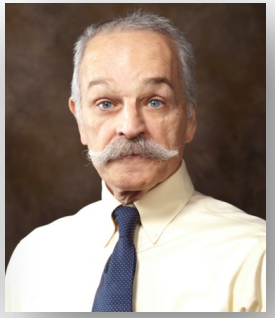
- Usage is up about 10 fold since review in April: 92,215 jobs on behalf of 6,932 users



Jetstream
http://jetstream-cloud.org/

Douglas Lab

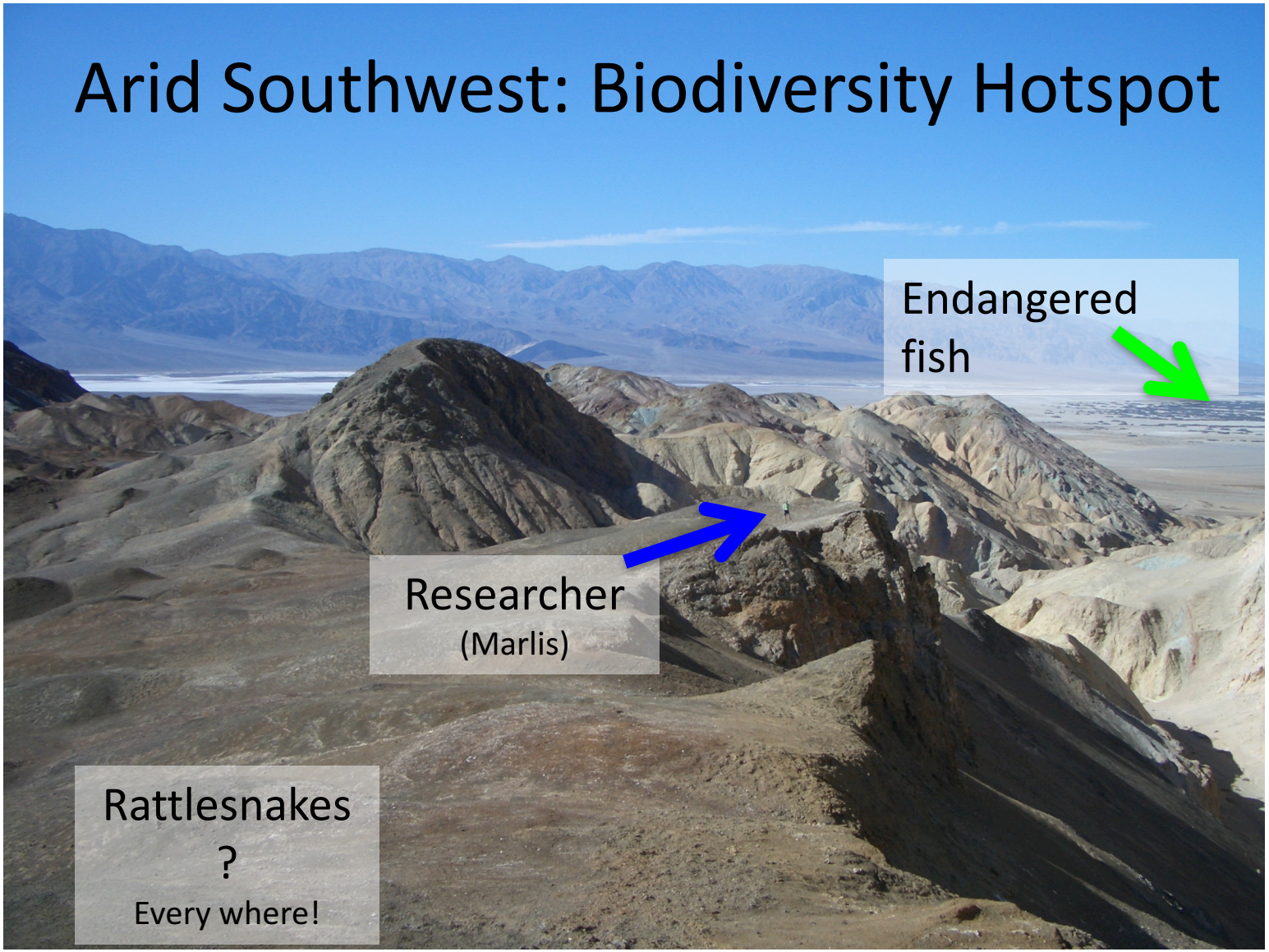
PIs / Advisors



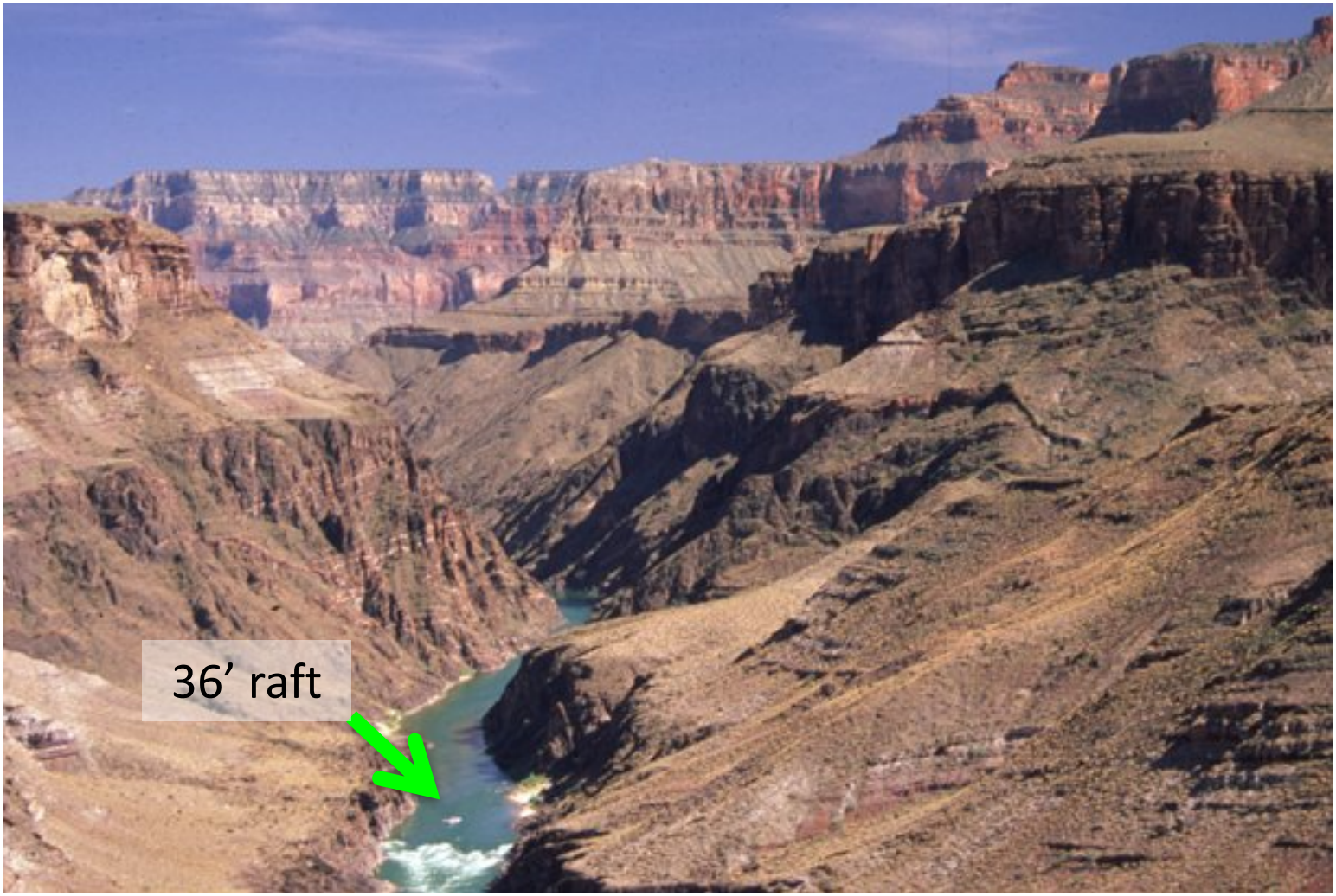
Collaborator

Postdoc

Graduate Students


Arid Southwest: Biodiversity Hotspot
Endangered fish
Researcher (Marlis)
Rattlesnakes ?
Every where!

Fieldwork: Grand Canyon



Western Fishes: Conservation

- Endemic species with adaptations to desert rivers
- Populations declining due to habitat alterations (= dams, stocking of predators, water diversions)
- Most such endemic species are rare, threatened or endangered

Ecosystem Change

Endangered

Invasive Species

Study Goals:

- Clarify biodiversity within Western Fishes
- Assess geographic patterns in genetic diversity
- Evaluate historic + current hybridization to test for gene flow between species

Slide courtesy Douglas Lab, University of Arkansas. Not to be reused without permission from the Douglas Lab

Data: ddRAD (double-digest Restriction-Associated-DNA

- Genomic Methods:
  - Extract genomic DNA from snake blood or fish fins
  - Cut genome in small fragments (~ 500 base pair)
  - Subsample ~40-60K fragments to reduce genome size
  - Sequence fragments with Next-Generation-Sequencing
  - 1 Illumina HiSeq lane: generates ~140 million reads

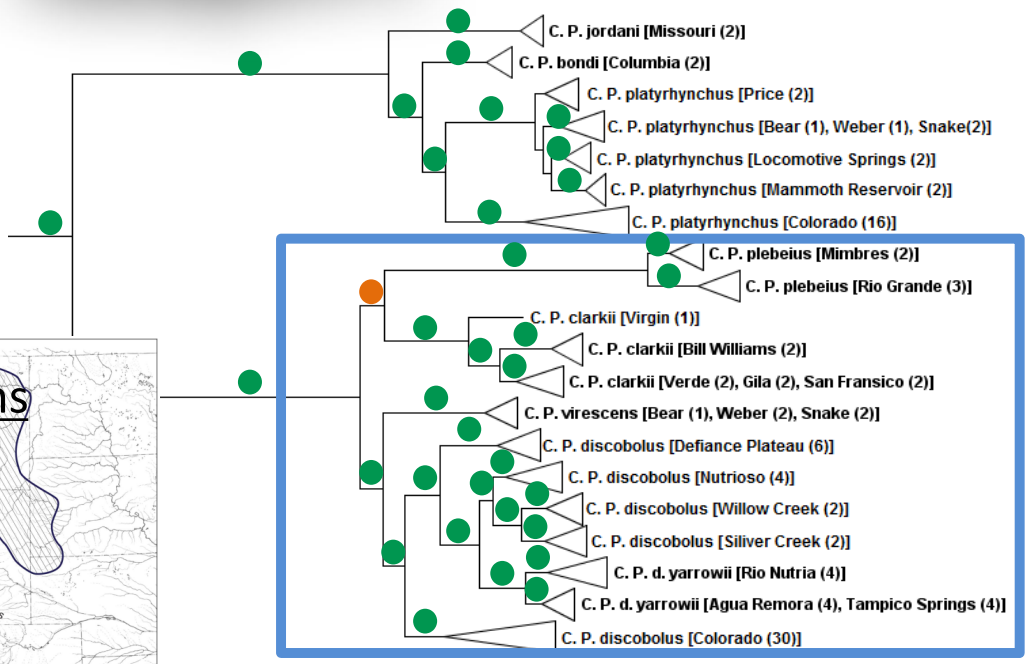
- ddRAD Project:
  - 6-8 HiSeq lanes / project
  - 400-800 samples of fish or snakes

} = large data sets

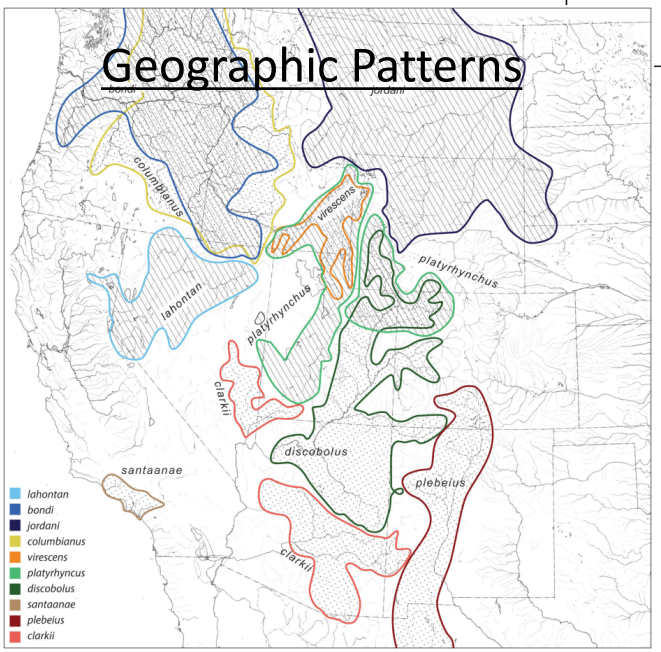
Western Fishes

Suckers (*Pantosteus*)

Phylogenomic Tree



Geographic Patterns

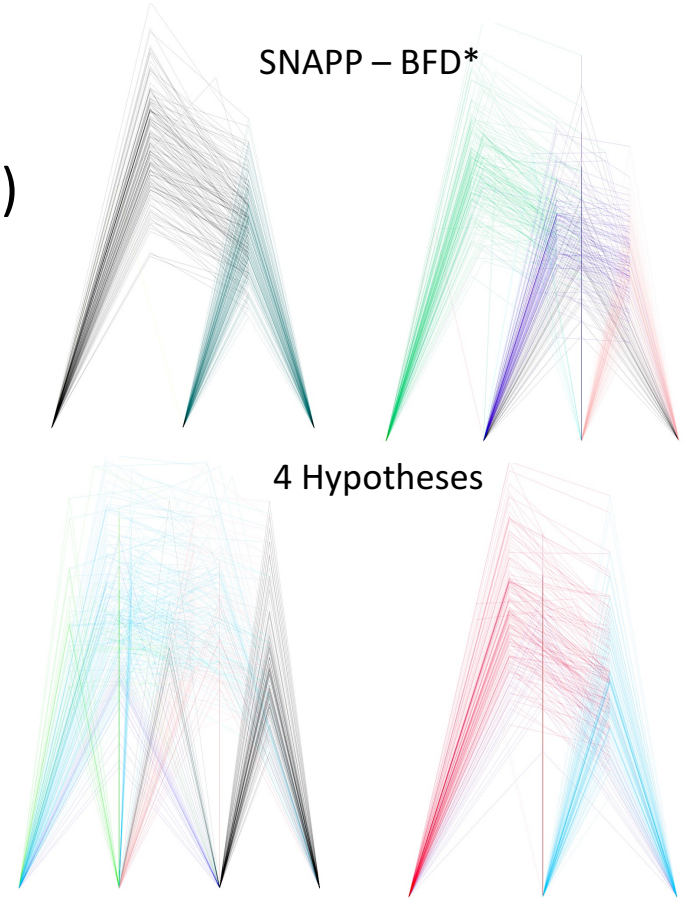


Species Tree: 152 samples
20,325 loci
206,736 SNPs
15x average coverage

Western Fishes: Analyses on Jetstream

Bayes Factor Delimitation (BFD)
(Species delimitation)

- Compares phylogenetic hypotheses
  - Incorporates thousands of loci
  - Integrates over gene tree topologies
- Computationally intensive
  - MCMC algorithm
  - Millions of iterations



Jetstream Results - BFD

Model	Species	Marginal L	BF	2xLN (BF)	Rank	Mean ESS
virgin+lcr+colorado	2	-3967.26	-	-	5	748.62
virgin, lcr+colorado	3	-3746.59	220.67	10.79	2	707.51
lcr, virgin+colorado	3	-3819.61	147.65	9.98	4	696.74
colorado, virgin+lcr	3	-3778.83	188.43	10.47	3	697.3
colorado, virgin, lcr	**4**	**-3642.89**	**324.37**	**11.56**	**1**	**702.22**

BFD selects model 5: 3 distinct species

Project 2: Phylogenomics of the Western Rattlesnake (*Crotalus viridis*) species complex

Western Rattlesnake: Analyses on Jetstream

Genomic fragments aligned + clustered (pyRAD)

Tested 3 clustering thresholds (0.85%; 0.90%; 0.95%)

- Lower values = more missing data allowed

Phylogenetic Tree constructed (RAxML)

- Compared topologies (patterns) and support values (statistical confidence) across different trees
- NB: RAxML is 4th generation descendant of Joe Felsenstein's DNAml program (IU was responsible for 3rd generation in this family tree of codes)

Key points

Field researchers with insufficient computational power available for their research needs

Data sitting around unanalyzed; Jetstream sped up research

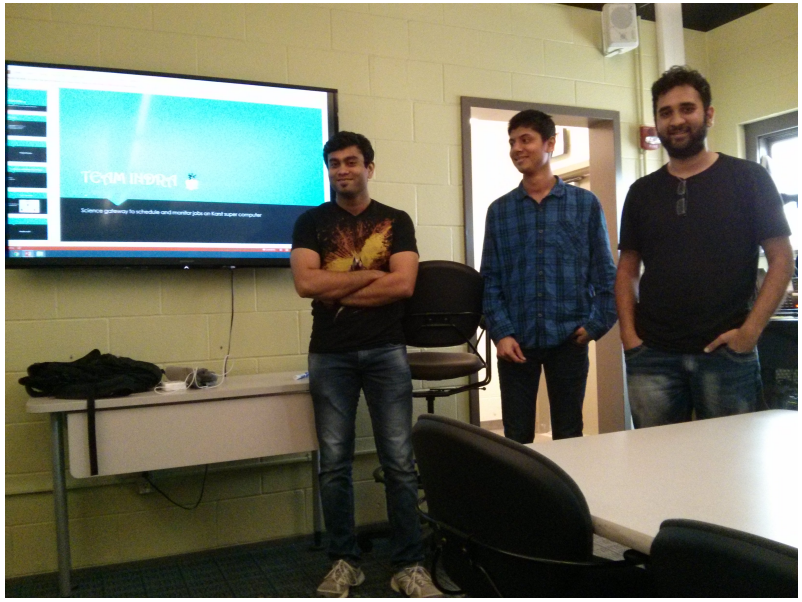
Relationships mediated by XSEDE Campus Champion

Meritorious science

Important broader impacts

Other projects ongoing

- Volker – eusocial insect evolution
- Cornell U (Lifka et al) – project Aristotle – have moved VMs among Redcloud, Jetstream, Amazon, Google, and Azure
- Pierce – teaching computer science
- Pestilli – brain science
- Numerous classes (some large) using Jetstream



Students using Jetstream in a team project demonstration at the same time as the acceptance review.

Some work already done with Jetstream

Earth & Atmospheric science: IRIS, UNAVCO, Unidata

- IRIS & UNAVCO planning to move data distribution infrastruyctre to Jetstream
  - IRIS: Goal is to deliver large amounts of curated and created seismology/geology data sets
  - UNAVCO: Plans to move to Jetstream with renewal
  - Both proposals depend on utilizing Wrangler for near-line dataset storage and delivery.
- Unidata: Goal is to deliver large amounts of real time meteorological data as well as to provide a platform for educational services using Docker (Mohan Ramamuthy allocation: Atmospheric Science in the Cloud: Enabling Data-Proximate Science)

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

GenApp Gateway

- Developed by Dr. Emre Brookes from the UTSA Health Center
- Provides a framework for processing all manner of jobs utilizing modules and wrappers to simplify software use and automation*
- Currently in alpha testing on Jetstream running NAMD jobs as a gateway using true elastic computing techniques to spin up multiple instances based on computing need

* https://figshare.com/articles/Creating_Science_Gateways_with_GenApp/4495865

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Atlas/OSG

- Atlas was one of the first research allocation holders on Jetstream
- Working on spinning up Jetstream instances on demand for additional data processing of LHC data
- OSG is currently planning an OSG software image that can be spun up on demand (based on thresholds defined by Jetstream admins) to become part of the OSG processing pool when Jetstream has cycles to spare

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Other top allocations on Jetstream

- **Inter-cloud Bursting: Decreasing Time-to-Science with a Multi-Stack Cloud Federation** – Adam Brazier, Cornell
- **Science and Engineering Applications Grid (SEAGrid): A Gateway for Simulation of Molecular and Material Structures and Dynamics** – Sudhakar Pamidighantam, Indiana University
- **Characterizing Extrasolar Planets with Jetstream: De-biasing High-Contrast Photometry** – Jared Males, University of Arizona (this is a project where work started on Chameleon and is migrating to Jetstream)
- **The Galaxy XSEDE Gateway** – James Taylor, Johns Hopkins University
- **A RADICAL Use of XSEDE: Abstractions Driven Cyberinfrastructure for XSEDE Resources** – Shantenu Jha

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream User Survey 2016 – Preliminary Results

Craig Stewart, Ph.D

Principal Investigator for NSF-funded Jetstream Cloud System

Jetstream User Survey

- Conducted December 2016 - January 2017
- All users (900+) included in survey sample
- 71 respondents in total
- Over 81% had used Jetstream since it went into production on June 1, 2016

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Satisfaction with Jetstream Services

Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, with 1 being “extremely dissatisfied” and 5 being “extremely satisfied.” If you have no basis for rating your satisfaction, please select “Not applicable.”		
Service evaluated	**Mean Satisfaction +/- 95% CI**	Lower, Upper Bound
Availability of VM images to solve my problems	4.02 +/- .23	3.79, 4.25
Speed (responsiveness) of Jetstream	4.02 +/- 0.19	3.83, 4.21
Documentation about Jetstream	3.84 +/- 0.22	3.62, 4.06
Jetstream Portal	4.09 +/- 0.17	3.92, 4.27
Speed of response to my questions via	4.55 +/- 0.15	4.40, 4.71
Quality of response to my questions via	4.60 +/- 0.14	4.46, 4.74
Speed of response to my questions via direct email to	4.64 +/- 0.15	4.50, 4.79
Quality of response to my questions via direct email to	4.64 +/- 0.15	4.50, 4.79
Overall performance of Jetstream	4.21 +/-0 .21	4.0, 4.42

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Importance of Jetstream in Research and Education Activities

Please rate the importance of Jetstream to your research activities on a scale of 1-5, with 1 being "not important at all" and 5 being "essential." If you have no basis for rating Jetstream's importance to your research activities, please select "Not applicable."

	Mean Importance +/- 95% CI	Lower, Upper Bound
Research Activities	3.72 +/- 0.28	3.44, 4.00
Education Activites	3.71, +/- 0.30	3.41, 4.02

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Comments about Jetstream – Positive feedback

- I believe Jetstream constitutes a major millstone in the advance of HPC systems for scientific computing. For basically the first time, cutting-edge computing resources are available to the long tail of scientific computing – disciplines which have until now been seen as out of scope for HPC applications. This is still early days for Jetstream, and the previously excluded researchers and disciplines are only just beginning to discover and experiment with Jetstream. I think if Jetstream persists, we will discover new opportunities and modes of commuting that drive innovation in a way that has been impossible on either the specialized and locked-down architectures of classic HPC, or the commercially-controlled public cloud of AWS and other private firms.
- I rely on access to the underlying OpenStack API, so I very much appreciate that the system is not limited to the Atmosphere layer! Thanks for providing the flexibility. I wish more attention and effort was invested in image management and curation. This is an area where I would like to and help in any way tat I can, since ready-made environment are essential to getting other researchers up and running quickly and effectively.
- Cloud systems such as Jetstream are essential to experiment with different collection architectures, data workflows, data management and archiving. This resource allows exploring combinations, integration, distribution and functionalities for data infrastructure. It is a great resource that would be expensive and would not allow such flexibilities.
- I think Jetstream is an amazing resource that was sorely missing from the HPC and HTC landscape. This amazing system should be expanded as it is a invaluable research tool for some, critical training ground for others, and gateway to bigger and more complex problems for all. I hope NSF will continue to make investments in Jetstream as it becomes an indispensable tool for the STEM research community and a valuable training tool for students. Kudos to all involved! The support is excellent, the design is superb, and the usefulness is extensive.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Comments about Jetstream – Constructive Critique

- Jetstream is [an] extremely valuable resource; however, documentation is almost non-existent. During the last semester, it was crashing multiple times without providing adequate error messages. I am looking forward to video tutorials that would depict some best practices of working with Jetstream from the research perspective. I would be specifically interested in learning how to move existing Oracle database[s], and to employ any data visualization features (if any) for data analysis.
- I understand that more documentation is coming every day. I marked [documentation] as something less than SuperDuperOutstanding. You are literally writing the book on this new resource and I want to cheer you along in that endeavor. I am especially interested in documented examples of cloudy techniques that others are using,
- VMs are not stable. Many VM images will not deploy. Shutting down, restarting, suspending, and resuming tasks usually ends up in hung VMs.
- I would suggest having a smaller number of VMs, but better support... For example, some VMs include a MatLab installation, but the executable path is not set, so the user has to go through the filesystem looking for the binaries. Why not link the installed programs? It might be helpful to include a README file with each VM that helps guide the user on how to install specific programs, etc.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Where can I get help?

Wiki / Documentation: http://wiki.jetstream-cloud.org

User guides: https://portal.xsede.org/user-guides

XSEDE KB: https://portal.xsede.org/knowledge-base

Email: help@xsede.org

Campus Champions: https://www.xsede.org/campus-champions

Training Videos / Virtual Workshops (TBD)

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream Partners

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Questions?

Project website: http://jetstream-cloud.org/

Project email: help@jetstream-cloud.org Direct email: jeremy@iu.edu

License Terms

- [CITATION HERE]
- Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- This research was supported in part by the Indiana University Pervasive Technology Institute, which was established with the assistance of a major award from the Lilly Endowment, Inc. Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc.
- Items indicated with a © are under copyright and used here with permission. Such items may not be reused without permission from the holder of copyright except where license terms noted on a slide permit reuse.
- Except where otherwise noted, contents of this presentation are copyright 2015 by the Trustees of Indiana University.
- This document is released under the Creative Commons Attribution 3.0 Unported license (http://creativecommons.org/licenses/by/3.0/). This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

CYVERSE™

Atmosphere

Edwin Skidmore
edwin@cyverse.org

CYVERSE™

Overview of CyVerse

Vision:

Transforming science through data-driven discovery

Mission:

Design, develop, deploy, and expand a national cyberinfrastructure for life science research, and train scientists

Funding:

National Science Foundation

Usage:

More than 41K users, 2.1 PBs of data, and hundreds of publications, courses, and discoveries

http://www.cyverse.org/

CYVERSE™

Atmosphere background

- Atmosphere first launched in 2011 by iPlant Collaborative (former moniker of CyVerse)
- Vision: A software enabling reproducible, collaborative scientific discovery with the tools and data users want using any cloud
- Multi-cloud orchestration technology for cloud resources
- Focuses on a simplified user experience of cloud computing for scientists and engineers, not the technology

CYVERSE™

Jetstream Atmosphere

- Integration with OpenStack Kilo/Liberty/Mitaka
- Integration with Globus
- Integration and modeling XSEDE allocation, including reporting
- Adapting to Jetstream-specific requirements (e.g. networking, configuration, etc)
- At the beginning of Jetstream, the code was forked; now, both Jetstream and CyVerse Atmosphere is the same code.

https://github.com/cyverse/atmosphere

CYVERSE™

Roadmap 2017 Q1

- Dynamic tool installation: Installation and configuration of tools on running instances (versus imaging baking)
- Image owner enhancements
  - Better restrictions and controls
  - Better metrics and reporting
- Simple DOI association
- Project sharing
  - Users in the same XSEDE allocation
  - Workshops and courses
- More transparency into OpenStack layer
- OpenStack Newton integration

CYVERSE™

Roadmap 2017 Q2

- ElasticSearch integration
  - Images (e.g. indexing packages and tools)
  - Volumes (e.g. indexing user data)
  - Instances (e.g. indexing logs for performance)
  - Whole system (e.g. indexing for system health and metrics)
- OnDemand Virtual Clusters
  - Ability to launch multiple instances
  - Tools to configure instances using relative ordering or user-defined variables (e.g. master/slaves, workers/coordinators)

CYVERSE™

Roadmap 2017 Q3

- DOI provider services integration
- Volume enhancements, including
  - Support for bootable volumes
  - Volume “Copying” to other users
  - Volume backups
- Jupyter integration
- Advanced networking management
  - IP reservations
  - DNS assignment (if enabled)
- OpenStack Ocata Integration

CYVERSE™

Roadmap 2017 Q4

- Container integration
- Instance state workflows: ability for site operators to define rules on instance states and taking actions on behalf of the user (e.g. if an instance has been idle too long, suspend the instance; then shelve an instance after being suspended for a length of time)
- User-defined workflows: ability for users to execute tools on a vm non-interactively

CYVERSE™

Questions?

CYVERSE™

CYVERSE™

CyVerse Collaboration

CYVERSE™

Goals

- Identify workflows or tools that leverage Jetstream's unique capabilities
- Provide a pathway for CyVerse users to Jetstream who may need
  - larger scale cloud resources
  - larger allocations (versus CyVerse's periodic allocation model)
  - CyVerse's cloud users are familiar with Jetstream's UI
- Deeper integration with CyVerse's other services
  - Agave API
  - Discovery Environment

CYVERSE™

WQ-MAKER Workflow: Background

- MAKER is a flexible and scalable genome annotation pipeline (Mark Yandell Lab from Utah)
  - De novo genome annotation
  - Updating existing genome annotation
  - Combining evidence with genome
- Limitations of MAKER
  - Installation of MAKER is challenging and complex
  - MAKER runs are not time efficient
  - On CyVerse, most users can only run a few instances using Maker-P

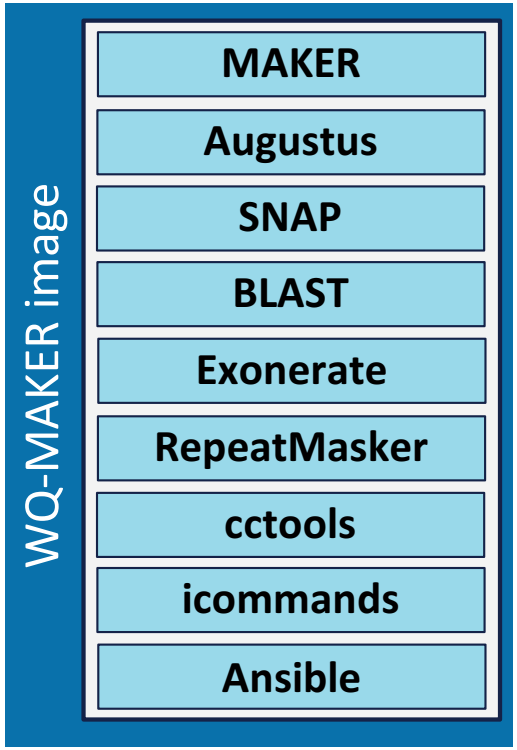
CYVERSE™

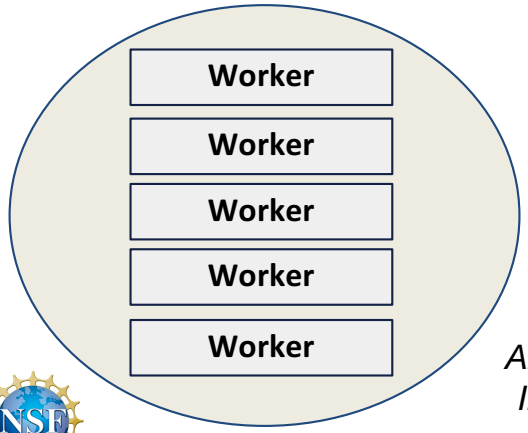
WQ-MAKER Workflow: Background

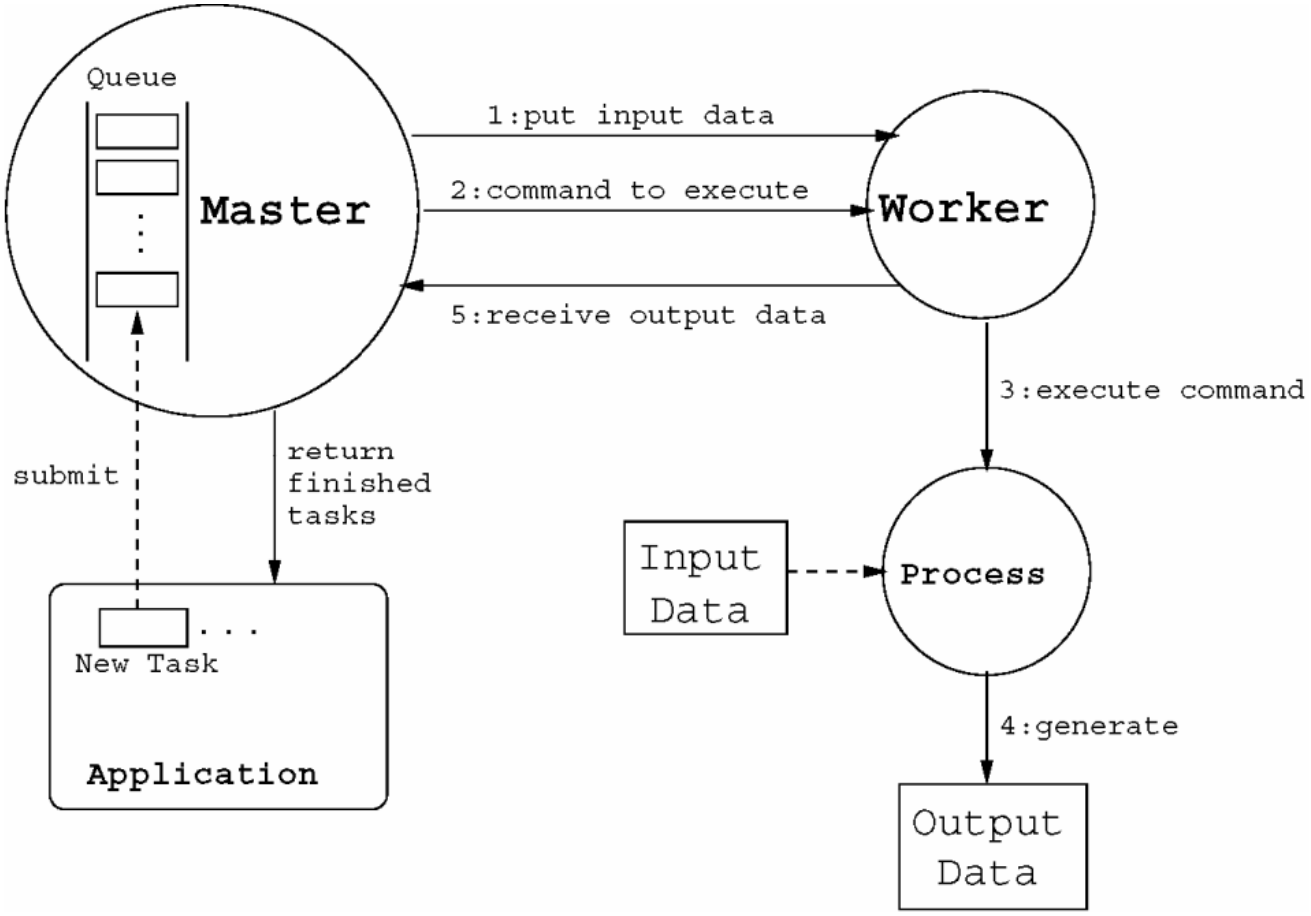
- WQ-MAKER is a modified MAKER annotation pipeline capable of being run on distributed computing resources using Work Queue (Doug Thain Lab from U of Notre Dame)
- Initially build to run on Amazon
- Upendra Devisetty (CyVerse) adapted WQ-MAKER for Jetstream
  - Developed the scripts to configure and connect master/slave nodes
  - Pre/post-processing tools
  - Visualization
  - Allows CyVerse users to pull large data from CyVerse Data Store
- Future work:
  - Integration with CyVerse Discovery Environment -> Jetstream
  - Manuscript forthcoming

CYVERSE™

WQ-MAKER Workflow: Jetstream







Scaling up genome annotation using MAKER and work queue

Andrew Thrasher, Zachary Musgrave, Brian Kachmarck, Douglas Thain, and Scott Emrich
International Journal of Bioinformatics Research and Applications 2014 10:4-5, 447-460

WQ-MAKER Workflow: Users

- In January 2017, WQ-MAKER on Jetstream was unveiled at Plant and Animal Genome, one of the largest international genomics conferences
- One user already has an allocation to begin her work; CyVerse working with two other researchers to get them a starter allocation
- On February 24, Upendra Devisetty will host a webinar on using WQ-MAKER in Jetstream: http://www.cyverse.org/blog/events/webinar-wq-maker-flexible-scalable-genome-annotation-pipeline-jestream-cloud

CYVERSE™

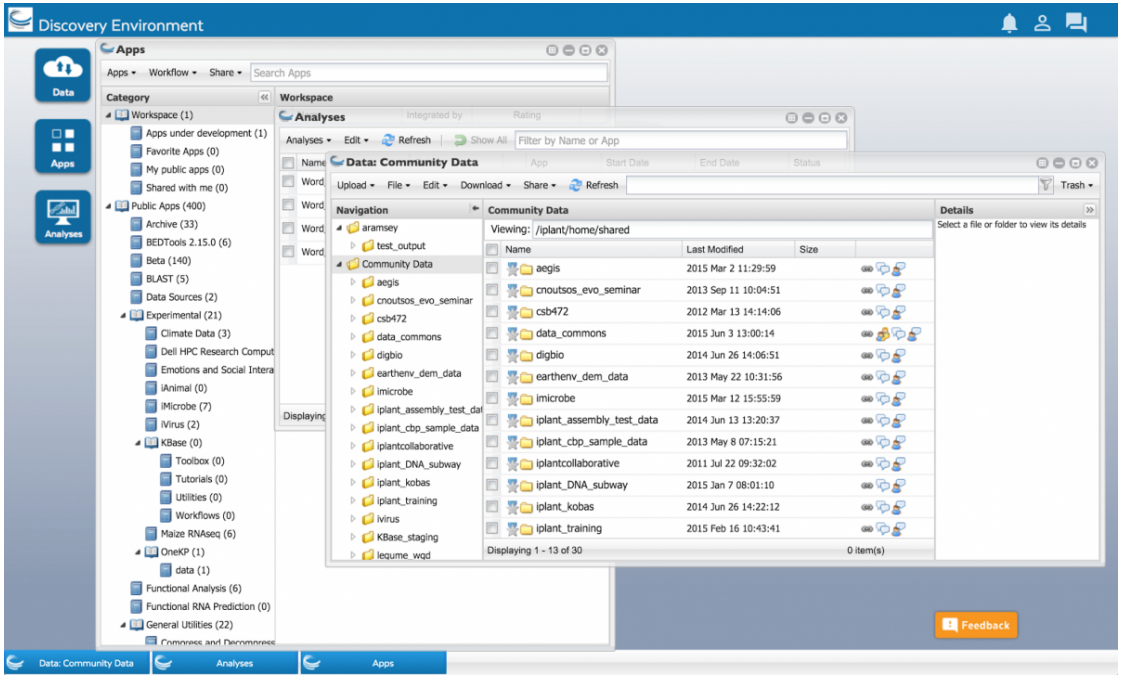
Gene Expression Matrices (GEMs) Workflow

- Adapting a popular workflow originally used on Open Science Grid (OSG): https://github.com/feltus/OSG-GEM
- Uses Pegasus Workflow Management System
- Currently, workers must be statically added to workflow
- Future work will include dynamic provisioning and automatically added to workflow
- Led by Upendra Devisetty

CYVERSE™

Integration with CyVerse Discovery Environment

- CyVerse's web interface that provides convenience access to data management and computation tools
- Abstracts the complexity of integrating with storage and compute infrastructure
- Users will be able to use their Jetstream allocation in the DE



Thank you

CYVERSE™

Jetstream Software Roadmap

Major Challenge is reconciling three IAM/accounting/reporting models

	Who	Type	Unit	Storage	Period	Overage	Access
Cyverse	Anyone	Personal	AU	No	Month	Yes	Role
TACC	PI	Group	SU	Yes	Quarter	No	Person
XSEDE	PI*	Group	SU**	Yes	Year	No	Person

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Accounting

- Revoking of access rights
- Culling policy for paused or abandoned assets
- Renewal and supplement support
- 30 day grace period for access
- Management tools for Allocation Owners
- Overage support
- Manage and report storage allocations
- Enhanced integrity checking for XSEDE reporting

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Easy Button

It's still too cumbersome for novices to get access to Jetstream

Let any user with active XSEDE User Portal account use a bit of Jetstream

- Get an XSEDE account
- Sign in to XSEDE User Portal
- Click "Trial Jetstream Access" button
- Get access to Jetstream in about ~~30 minutes~~ 4 hours

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Implementing the Easy Button

This is coming along, though delayed

Implementing a restricted “role” in Atmosphere for EZB users

- One m.small VM
- 1 IP address
- 1x 10GB volume
- 1 snapshot

Resolving behavior when a user joins a real allocation
Providing better resource dashboard for EZB users

Object Storage

Implement S3-compatible object storage

- Develop user interface in Atmosphere
- Implement utility feature set (i.e. bucket-name = DNS record)
- Resolve storage quota and accounting

Openstack providers have this turned on now. May leave it as a power user function for quite some time.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Custom DNS

We provide public IP addresses, but a lot of cloud's flexibility comes in being able to manage DNS.

Static IP -> Update DNS ***or*** **Static DNS -> Update IP**

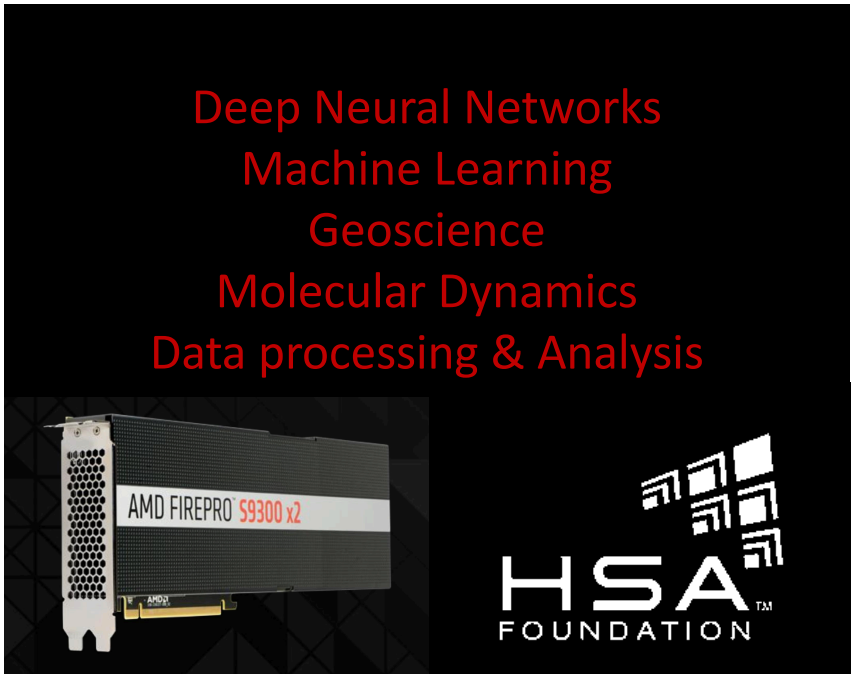
Our IP pool is limited

Most users DON'T know how to manage DNS

Specify **hostname**.jetstream-cloud.org (or other TLD) at boot

Work is beginning to build this off **OpenStack Designate**

Hardware Expansion Plans



Unrestricted $75,000 Gift from **Heterogeneous System Architecture (HSA) Foundation**

- 10-12 nodes Intel Haswell nodes
- 1-2 TB SSD ephemeral disk
- 256 GB RAM
- Dual AMD FirePro S9300x2

1 TB/sec memory bandwidth
13.9 TFLOPs single precision

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Globus features

Lee Liming (lliming@uchicago.edu)
Computation Institute
University of Chicago

funded by the National Science Foundation
Award #ACI-1445604

Globus features on Jetstream

Basic features (available now)

- Researchers must be able to use XSEDE identities to login to Jetstream's Web UI.
- Researchers should also be able to use campus (InCommon) identities to login to Jetstream's Web UI.
- Researchers must be able to use Globus to move data into/out of Jetstream Virtual Machines.

Advanced features

- If Jetstream allows researchers to **import/export VM images**, then researchers should be able to use Globus to do that.
- Researchers should be able to transfer **volumes** to/from Jetstream easily and as efficiently as the networks allow.
- When Jetstream offers S3-style storage services, they should be able to use Globus to **transfer S3 buckets** to/from Jetstream.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Moving data to/from Jetstream VMs

- Today, researchers can do this by installing the Globus Connect Personal (GCP) agent in their VM, as they would do on any other system they use.
  - Performance isn't as good as it would be with Globus Connect Server (GCS).
  - Installation is quite easy, but not as easy as pushing a button or having it be there automatically.
- Our plan for 2017 is to **pre-install** GCP on most/all Jetstream featured images and **pre-register** a Globus endpoint.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Importing/exporting VMs using Globus

- This would be part of Jetstream's image publication feature, and would help in cases where the researcher chooses to publish in an archive that supports Globus and/or Globus Publication.
  - At the moment, there are no mainstream archives like this.
  - Image publication isn't a high-demand feature in Jetstream (yet).
- This feature is not in our 2017 plans, for the reasons above.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Volume transfers

- Jetstream allows researchers to attach storage volumes to VMs up to ½ TB (500 GB) in size.
  - Given the storage limitations, the current GCP transfer performance is reasonable.
  - It would be better to offer GCS-scale performance, especially if storage sizes get larger.
  - It *might* be desirable to enable users to access storage volumes without activating a VM (using a common Jetstream endpoint).
- This also is not in our 2017 plans.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

S3 storage and Globus

- Jetstream has not (yet) provided S3 storage services, but the OpenStack technical team is close to being ready to do so.
  - We will need a UI for creating/managing S3 storage buckets.
  - We will (eventually) need a transfer mechanism that can move S3 buckets between Jetstream and other systems.
- In 2017 (Q1 and Q2), our plan is to install the Globus S3 connector software on Jetstream data transfer nodes and offer it as the transfer mechanism. UC (Globus) and UA (Atmosphere) will collaborate on the UI.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream Partner Organizations

Initial construction partners

Indiana University

The University of Texas at Austin

TACC

The University of Arizona

The University of Chicago

globus

Johns Hopkins University

Management & Operations partners

The University of Texas at San Antonio

Penn State

Cornell University

Application / community lead partners

University of Hawai'i

University of Arkansas at Pine Bluff

NSIDC National Snow and Ice Data Center

UNC The Odum Institute

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Questions?

Project website: http://jetstream-cloud.org/
Project email: help@jetstream-cloud.org Direct email: YourEmailHere@addr.org

License Terms

- *[Citation if needed]*
- Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- This research was supported in part by the Indiana University Pervasive Technology Institute, which was established with the assistance of a major award from the Lilly Endowment, Inc. Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc.
- Items indicated with a © are under copyright and used here with permission. Such items may not be reused without permission from the holder of copyright except where license terms noted on a slide permit reuse.
- Except where otherwise noted, contents of this presentation are copyright 2015 by the Trustees of Indiana University.
- This document is released under the Creative Commons Attribution 3.0 Unported license (http://creativecommons.org/licenses/by/3.0/). This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream Support for Science Gateways

Marlon Pierce

What Is a Science Gateway?

Science gateways are user interfaces and supporting middleware that provide access to scientific applications and data to communities of users.

Software as a Service for science

Infrastructure for Science Gateways

Use Case	Description
UC1: Gateway Hosting	Gateways are Web portals. They need Web servers, databases, Web programming languages, etc. May be standalone, use Galaxy, or use a PaaS
UC2: Gateway Platform as a Service (PaaS) hosting	The gateway PaaS middleware itself needs to be hosted. Examples include SciGaP.org, HUBzero, and the Agave Project.
UC3: Gateway Platform as a Service Research	Gateway PaaS systems are also laboratories for pragmatic distributed computing research and development.
UC4: Computing power for running scientific applications	XSEDE gateways use the same job submission queues as regular users; gateway accounts are mapped to “community” user accounts. Frequently gateways need more direct control over resources than this.
UC5: Large scale storage co-located with Web infrastructure	Gateways provide access to community data sets, not just applications. They need a way to mount large data sets and make them available through Web interfaces and services.

Gateway Related Allocations by the Numbers

- Over 30 science gateway-related allocations
  - ~14% of all active allocations
- >14M XSEDE SUs awarded to gateway related projects
  - >42% of all awarded SUs
- 6 projects receiving XSEDE ECSS science gateway support
- 2 projects receiving Science Gateways Community Institute Extended Developer Support services

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream Plans for Science Gateways

XSEDE Gateway Hosting

- Position Jetstream as the resource for hosting gateways, web servers, Web services, databases, and similar community resources.
  - Replaces IU's Quarry
- Will be finalized at the next XSEDE quarterly meeting (March)
- This will require an audit of current Quarry usage.

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

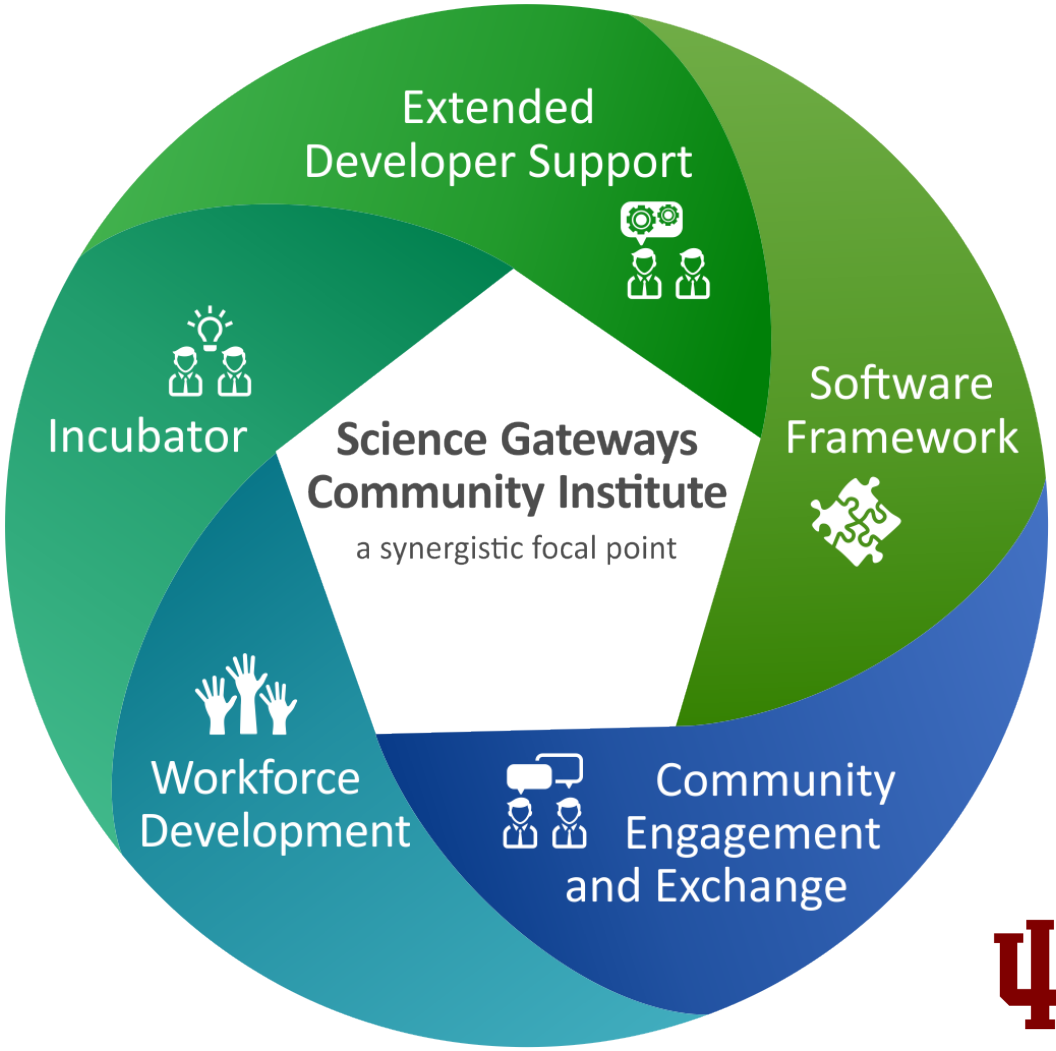
Quarry Gateway Hosting Transition

- IU's Quarry has 175 registered VMs
  - 60 VMs requested via XSEDE allocation process
  - 25 VMs for XSEDE services
  - 90 VMs from internal IU requests
  - Many projects have multiple VMs
- Wide range of usage
  - Gateways,Web servers, Wikis, RPM repos, ...
  - login.xsede.org and other XSEDE services
  - Testbeds
- At least15-20 gateways will need to be moved to Jetstream or an alternative location
  - 12 are related to the IU-led SciGaP.org collaboration
- Determining

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

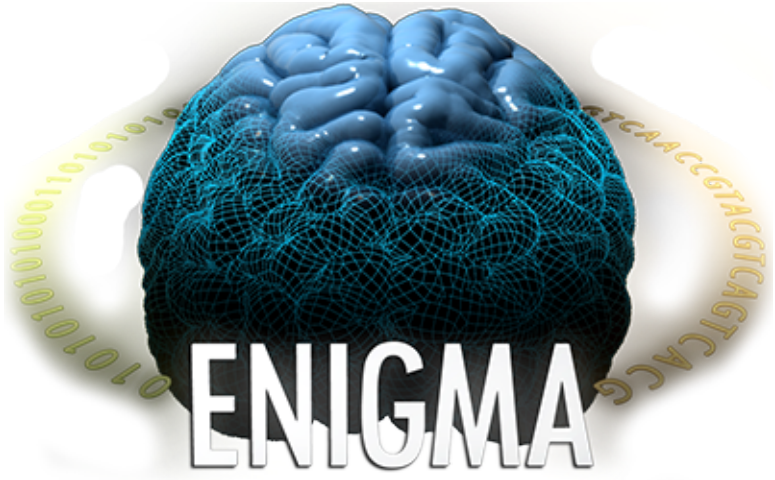
Leveraging Science Gateways Community Institute Services



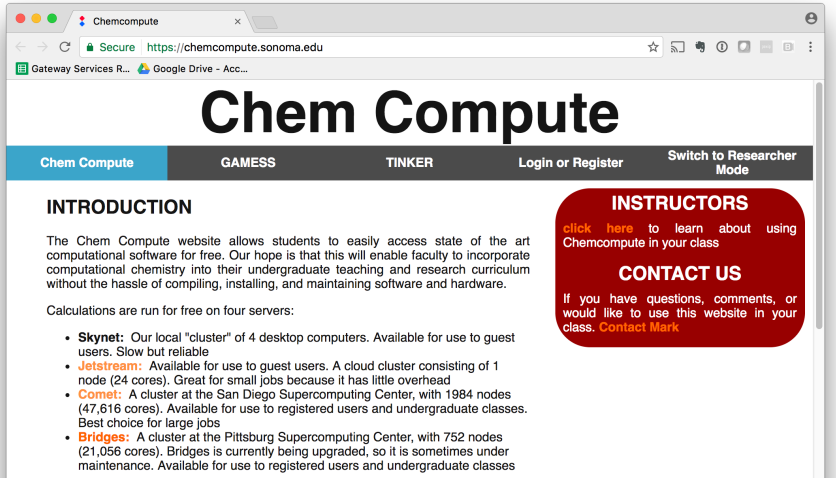
The SGCI is now an XSEDE Level 2 Service Provider

Additional Slides

Gateway Hosting Examples



The ENIGMA Bipolar Disorder Brain Age project aims to understand brain aging in bipolar disorder (BD) by creating a large database of existing measures of brain size.



Chem Compute provides easy access for undergraduate chemistry students to use computational / quantum chemistry packages for their classes or research.

Gateway Platform as a Service Hosting Example

Jetstream
http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604

Gateway PaaS Research Example

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Gateway Execution Support Examples

Gateway Data Access Example

These are the UNAVCO and IRIS projects.

jetstream

Looking forward to new technology

2 Part discussion:

1) Future Thoughts about cloud. Paul Rad, Ph.D., Co-founder and Chief Research Officer. Open Cloud Institute (OCI), University of Texas and San Antonio, paul.rad@utsa.edu

2) If we had more money (or XSEDE were to spend more money on us). Craig Stewart

funded by the National Science Foundation
Award #ACI-1445604

Cloud Trends for 2017 – Related to Jetstream

The Transition to Multi-Clouds Accelerates. And Infrastructure as a Service (IaaS) and Platform (PaaS) as a Service Convergence

- The distinctions between infrastructure and platform are evaporating, leaving these terms to collapse into a single concept
- Research workloads will be distributed among public clouds (Amazon, Google, Microsoft), Managed Science Clouds (Jetstream), and Campus clouds

Advanced Analytics becomes more accessible.

- Data gravity will push scientific computing workloads to where data lives and Cloud Data Warehouses will be the popular data destinations for research computing

Edge Computing extends the power of cloud to connected Internet of Things (IoTs) for distributed data processing

Containerization is emerging as a platform for distributed multi-cloud computing

Serverless (AWS Lambda, Azure Functions, Google Cloud Functions, Jupyter Notebook) is emerging as a platform for browser based thin desktop computing

Education

- Talent and expertise is the biggest challenge
- Data Literacy becomes a fundamental skill of the future

It has been straightforward for really experienced users to integrate new tools. For example..

And if we had more money…. From the NSF or from XSEDE

Supplements to Jetstream award

- Additional outreach staff (pending)
- Funding to SUNY Binghamton for supporting orchestrated use of VMs on Jetstream – in prep
- More work related to Atmosphere and API?
- More work related to storage, including Wrangler integration

And if we were to get more money or attention from XSEDE

- National VM library function
- More ECSS support for software inside VMs
- More support for inside-the-VM performance analysis

jetstream

Outreach, Education and Training Plans for 2017
3 Part presentation:

1) Therese Miller, Program Director, Collaboration, Engagement and Interoperability – summary of planned events

2) Susan Mehringer slides on Cornell outreach activities

3) Paul Rad slides on outreach from UTSA

funded by the National Science Foundation
Award #ACI-1445604

Jetstream EOT Plan 2017

- Jan 9-12 MiniCourse in Statistics, Brandeis University,
- Jan 14-18 Plant and Animal Genome Conference
- Jan 17-20 UTSA Technical Training for Faculty and Staff
- Feb. 5 Galaxy Australasia Meeting 2017, Melbourne, Australia
- Feb. 20 AAAS Conference, Jetstream Poster
- Feb 24 WEBINAR: WQ-MAKER: A flexible, scalable genome annotation pipeline on Jetstream Cloud (Online)
- Feb 27-Mar 3 SIAM CSE17 Broader Engagement Conference
- Mar/Apr Texas A&M, Research talk and tutorial for faculty and staff
- Mar 27-30 DellHPC, Austin, TX
- May14 IEEE CCGrid 2017, Madrid, Spain
- June 4-8 American Society for Mass Spectrometry Conference

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Jetstream EOT Plan 2017 (cont'd)

- July 9-13 PEARC17 Annual Conference (workshop??)
- July 24-27 Earth Science Information Partners Summer Meeting
- Sept 20-24 Organization of Biological Field Stations Annual Meeting
- Nov. 12-17 SC17 (workshop??)

Jetstream
http://jetstream-cloud.org/

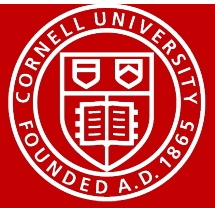
funded by the National Science Foundation
Award #ACI-1445604

EOT Targeted Audiences 2017

- Biological Field Stations (310)
- Engineering Programs
- Can anyone recommend good engineering conferences? (Is there an engineering equivalent to PAG in the biology community – Plant and Animal Genome conference?)

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604



Cornell University
Center for Advanced Computing

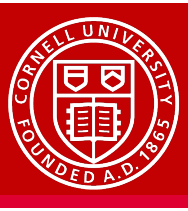
Cornell Virtual Workshop

Web-based training, comprised short audio or video clips, graphical simulations, examples, exercises, and quizzes, with full text discussion incorporating an HPC glossary

Provide education materials to the broader community, on demand

Four new Virtual Workshop modules will be developed, one per year

Modules will be reviewed and updated at least annually



Planned Topics:

PY1:

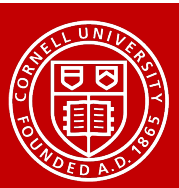
(a) Allocations walkthrough video with screenshots and key points
(b) Introduction to the Jetstream system and the Atmosphere cloud computing environment: using, creating, and archiving services and VMs

PY2: Using remote desktops to access and use Jetstream, XSEDE, and systems in the XD program

- Option to consider: change this to Engineering applications

PY3: Biology and earth science applications on Jetstream

PY4: Scientific replicability of computations and analyses: publishing, archiving, and curating your VM images for posterity



Currently in work:

Allocations short topic

- Walkthrough notes taken
- Video recording scheduled

Introduction to the Jetstream system and the Atmosphere cloud computing environment: using, creating, and archiving services & VMs

- Content materials gathered
- Outline written
- Content in work

Both planned to be complete Q1 2017

UT San Antonio outreach and pland

- Build Cloud Research and Education Platforms on Jetstream for Science and Engineering Communities
  - Cyber Security
  - Data Analytics and Machine Learning
  - Edge Analytics and Cyber Physical Systems

- Outreach
  - Promote Jetstream as a platform for K-12 Technology Literacy
  - Drive Collaboration with OpenStack Science User Community Globally

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

UTSA Open Cloud Institute (OCI) Jetstream Current Research and Education Activities

Conducted the 1st Jetstream workshop at UTSA Campus

- Introduced Jetstream environment to UTSA research community
- Identified early adopter faculties who will be leveraging Jetstream for their course projects (Spring 2017) and willing to share their results with the community
    1. Introduction to Cloud Computing (Computer Engineering)
    2. Intelligent Robotics (Control Engineering)
    3. Big Data with Machine Learning (Computer Engineering)

Building education platforms consist of Jetstream Images, hands on labs, and Jupyter notebooks on:

- OpenStack Cloud Computing – Hands on education video and education exercises to learn OpenStack services such as identity, compute, network, storage, dashboard, and etc.
- Big Data with Machine Learning – Hands on education exercises on graph analytics, linear algebra, probability, Information Theory, and deep learning models such as CNN and RNN
- Robot Operating Systems (ROS) - Hands on education exercises on robotic operating systems platform

UTSA Jetstream Upcoming Horizon

1) Will conduct the 2st Jetstream workshop at UTSA campus, Fall 2017 to share education and research conducted on Jetstream and assist the community to leverage Jetstream for their research and education

2) Develop value add capabilities on Jetstream to enable workload mobility

- Application containerization has emerged as a platform for workloads in scientific computing, in this effort UTSA will develop:

  - Develop new capabilities for integrating Docker containers with Jetstream in a loosely coupled architecture.

  - Investigate multi-clouds workload interoperability and Life Cycle Management practices and tools to enable workload mobility from locally managed campus infrastructure to shared public (Amazon, Microsoft, Google) or open managed science cloud i.e. Jetstream

UTSA Jetstream upcoming Horizon Continued

3) Develop advanced Testbed in collaboration with open communities to enable research and education on Jetstream

- **Cyber Security Testbed** - To handle the skyrocketing volume of malware and cyber exploits, we will build a cyber security testbed that provides a central place for conducting research and education on how to detect, diagnose, and remediate online attacks.
- **Data Analytics and Machine Learning Market Place** – As data becomes more open accessible and cloud technology enables easy sharing. Researchers can share live, interactive, executable environment to drive research collaboration

4) Outreach

- High School outreach to enable technology and data literacy
  - Build STEM curriculums on Jetstream with selected magnet schools in San Antonio
- Drive Collaboration with OpenStack Science User Community Globally → Boston Cloud Declaration
  - Provide greater interoperability between and among research clouds (global)